DOCUMENT RESUME

ED 208 771 HE 014 506

AUTHOR Davison, Mark L.; And Others
TITLE Higher Education and Cognitive-Social Development Project. Final Report. Project Report No. 8.
INSTITUTION Minnesota Univ., Minneapolis.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE 15 May 81
GRANT NIE-G-79-0021
NOTE 85p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Abstract Reasoning; *Cognitive Development; *College Attendance; College Freshmen; College Seniors; College Students; Comparative Analysis; Educational Benefits; Engineering; Graduate Students; Higher Education; Liberal Arts; *Majors (Students); *Moral Development; Moral Values; *Noncollege Bound Students; *Personality Development; Theories; Undergraduate Students

ABSTRACT

The impact of higher education on the cognitive and social development of students was studied in three experiments. A longitudinal comparison of moral reasoning development in the four years after high school among those who did and did not attend college was undertaken. Instruments used were the Reflective Judgment Questionnaire, the Concept Mastery Test, the Defining Issues Test, and the Sentence Completion Test of ego development. Over the first two years after high school, scores of both groups increased by about the same amount. In the second two years, the increase for the college attenders was greater than that for the nonattenders. Additionally, a cross-sectional comparison between undergraduates in liberal arts majors and undergraduates in engineering majors was conducted. On both a measure of verbal reasoning and a measure of complex reasoning about issues that have no simple right or wrong answer, seniors in both majors scored higher than freshmen. The senior-freshmen difference suggested a similar amount of growth in both kinds of majors. The third study assessed the impact of graduate education on cognitive and cognitive-social developmental measures. Advanced graduate students, first-year graduate students, and college graduates who had not entered graduate school were compared using five instruments measuring reflective judgment, moral reasoning development, ego development, verbal reasoning, and socioeconomic status. Age and selection appeared to explain differences on the measures between beginning and advanced students and between students and nonstudents. Theories and research relating to moral development, ego development and reflective judgment style are reviewed. Although methodological flaws in the research are acknowledged, among the conclusions are that people in college grow in moral reasoning (and at a faster rate than non-college attenders), and that college seniors do not differ significantly from freshmen in ego development. A bibliography is appended. (Author/SW)

ED208771

Final Report

Higher Education and Cognitive-Social Development Project

Grant No. NIE-G-79-0021

Mark L. Davison, James R. Rest,
Stephen J. Thoma and Elizabeth R. Welfel

University of Minnesota

Higher Education and Cognitive-social Development Project
Department of Social, Psychological, and Philosophical Foundations of Education

Funded by a grant from the National Institute of Education (NIE-G-79-0021)
Mark L. Davison, Principal Investigator

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

HE 011 506

Final Report

Higher Education and Cognitive-Social

Development Project

Grant No. NIE-G-79-0021

Mark L. Davison, James R. Rest,
Stephen J. Thoma and Elizabeth R. Welfel

University of Minnesota

Project Report #8

May 15, 1981

The research reported herein was performed pursuant to a grant with the National Institute of Education, U. S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U. S. Department of Education

National Institute of Education

Abstract

This project focused on the impact of higher education on the cognitive and social development of students. Study 1 was a longitudinal comparison of moral reasoning development in the four years after high school among those who did and did not go on to college. Over the first two years after high school, scores of both groups increased by about the same amount. In the second two years, the increase for the college attenders was greater than that for the nonattenders. Study 2 contained a cross-sectional comparison between undergraduates in liberal arts majors and undergraduates in engineering majors. On both a measure of verbal reasoning and a measure of complex reasoning about issues which have no simple right or wrong answer, seniors in both majors scored higher than freshman. The senior-freshman difference suggested a similar amount of growth in both kinds of majors. On measures of moral reasoning, personality development and two measures of cognitive development in Study 3, beginning and advanced graduate students differed by no more than two adult control groups. Results in Study 3 suggested that the observed differences between beginning and advanced graduate students were attributable to age and selection.

Final Report: Higher Education and
Cognitive-social Development Project

> What we should seek to impart in our colleges, therefore, is not so much learning itself as the spirit of learning. It consists in the power to distinguish good reasoning from bad, in the power to digest and interpret evidence, in a habit of catholic observation and a preference for a nonpartisan point of view, in an addiction to clear and logical processes of thought and yet an instinctive desire to interpret rather than stick to the letter of reasoning.
> (Woodrow Wilson, 1909)

There are at least two kinds of intellectual questions. The first we shall call intellectual puzzles. These are questions for which evidence and the rules of logic dictate a single, verifiable correct answer. An example is "$\sqrt{625}$ = ?" These are the kinds of questions with which most standardized tests deal.

But there is a second kind of intellectual question, which we shall call intellectual problems. These are questions which do not have a single, verifiable correct answer, because the evidence on the issue is incomplete or because it is contradictory. An example is "What proportion of America's energy needs can be met by solar energy in the 1980s?" There is evidence which can be brought to bear on this issue, estimates of how much energy American will need, estimates of how much of that need can be met by fossil fuel sources, and estimates of how much energy might be produced by solar technology. But the estimates made by different experts are inconsistent,

partly because the evidence which can be used in making those estimates is not sufficiently complete to dictate a single, logically verifiable, correct estimate.

Higher education courses deal both with intellectual problems and intellectual puzzles. For instance, elementary calculus courses, physics courses, and chemistry courses train students to solve equations or balance equations in a manner which will yield the correct answer. Courses in psychology, history, philosophy, and literature (to name just a few) force the individual to confront intellectual problems of historical interpretation, personal identity, social planning, and moral values which have no single, correct answer. We shall use the term reflective judgment to refer to an individual's own answer to an intellectual problems, because such answers are made after reflection on the available evidence, but they are judgments which go beyond that evidence. We shall use the term reflective judgment style to mean the way in which a person reasons about complex problems having no single correct answer.

The research in this project focused on the relationship between higher education and the way an individual reasons about intellectual problems. In addition, it focused on the relationship between higher education and three other cognitive variables which the literature suggests should be related to the development of reflective judgment style. Those three variables are verbal reasoning (the ability to solve verbal intellectual puzzles as tapped by conventional standardized tests), moral development, and ego development.

Three of the variables in this project, ego development, moral development, and reflective judgment, are what we shall call cognitive-social variables. By

cognitive-social, we mean that the variables involve the individual's understanding of the people in his/her social environment and his/her own relationship to those people. As will become clearer when we explain the theories on which our work is based, ego development, moral development, and reflective judgment are cognitive-social variables by this definition. As Table 2 shows, the individual's perception of interpersonal relationships is one of four aspects of ego development. Moral development involves the maturing of understanding about the principles which govern interpersonal transactions. And the highest levels of reflective judgment (see Table 1) presume an understanding of how other people think about complex issues.

In presenting the rest of the background for the project, we shall follow this outline. First we shall briefly describe the theories of moral development, ego development, and reflective judgment style out of which this research grew. We shall emphasize links between the theories, because these links have dictated our choice of research measures. Second, we shall briefly describe the most relevant empirical research in the areas of reflective judgment style, moral development, and ego development. This review must be highly selective, because the literature in these areas is so extensive.

Theories

Our work grows directly out of King (1977) and Kitchener's (1970) adaptation of William Perry's (1970) theory of intellectual and ethical development in the college years, Kohlberg's (1969, 1971) theory of moral development, and Loevinger's (1970) theory of ego development. All three

theories will be briefly described, along with the links between them.

All three of these theories are developmental stage theories. By basing our research on stage conceptions, we do not wish to espouse an overly simplistic theoretical conception, particularly a conception which describes an individual as reasoning at one and only one stage. According to our conception (Davison, 1977; Davison et al., 1978, 1980; Rest, 1979), the stage descriptors depict types of reasoning. Any given person will typically use different types of reasoning under different circumstances. Conversely in a given circumstance, different people will use different types of reasoning. While we can not characterize an individual by a number representing the level of reasoning always used by that person, the object of stage developmental measurement is to describe each individual by an index characterizing his/her "average" level of reasoning in a variety of circumstances. In our conception, the stage types are ordered in that (1) types of reasoning which are adjacent in the stage hierarchy are logically more similar than those which are nonadjacent and (2) people who frequently use a given stage, are more likely to use stages adjacent to that given stage than to use stages nonadjacent to the given stage (see Davison, 1977; Davison, et al., 1978; Davison & Robbins, Note 2; Rest, 1973).

King (1977) and Kitchener (1977) have jointly devised a more specific framework called Reflective Judgment which focuses on epistemological development. Influenced by the work of the philosophers Popper (1972), Kuhn (1970), and Dewey (1933), as well as psychologists Perry (1968), Boyd (1972), Riegel (1973), and Broughton (1975), King and Kitchener have argued persuasively that intellectual development continues beyond relativism.

Table 1

Description of Reasoning Characterizing Each Position

in Reflective Judgment

Position 1

Subjects use simple black and white, concrete, and categorical thinking. Knowledge is seen as absolute, and authorities are seen as the source of knowledge. Problems are solved simply by following the work of an authority, rules, tradition, or the norm. Judgement is seen as unnecessary, since alternatives are not acknowledged.

Position 2

Subjects perceive alternative views, but reject them without critical examination. They believe that the truth exists and that authorities usually know the truth. Their arguments are simple and frequently illogical. They may offer pieces of unrelated information as "evidence," and then choose a point of view on the basis of tradition or authority. They may also cite evidence which contradicts the view they endorse without acknowledging the contradiction.

Position 3

Subjects acknowledge the existence and temporary legitimacy of different views. Authority and knowledge become further separated, and they begin to see authorities as "biased" or arbitrary. With formerly-held absolutes no longer intact, the contingent nature of knowledge is confusing, as is decision-making. Everyone's view is seen as equally correct and/or equally biased. Decisions are based predominantly on personal belief or bias.

Position 4

Subjects acknowledge the lack of absolutes in some areas, but not in others. They begin to evaluate evidence, but do not understand that evidence entails a conclusion. They use both unsupported belief and evidence in decision-making. Subjects are often skeptical about the "truth value" of any evidence or an authority's opinion, and they deny that opinions or interpretations can be objectively evaluated.

Position 5

Subjects begin to understand that knowledge is embedded in a context and that a frame of reference is important for understanding a point of view. Authorities begin to be seen as experts who have reasoned to a point of view which may or may not be valid. Subjects evaluate evidence on several sides of issues and from several perspectives. They present a balanced view, but they do not integrate evidence into their own view.

Position 6

Subjects see knowledge as embedded in a context and are beginning to understand that not all points of view are equally correct. They can analyze evidence for alternative points of view and may argue on the basis of evidence that one is more likely. They do not synthesize the evidence into a view of their own, however. Usually, they rely on the synthesis of others (e.g., experts), whose views are also seen as open to evaluation.

Position 7

Subjects present an examined point of view which they endorse. It is based on an integration and logical evaluation of evidence, the opinions of experts, as well as reasoned conjecture about "what appears to be true." They are, however, willing to acknowledge that their views may be falsified and may need to be reformulated in light of additional evidence at a future time. Their point of view is presented as being probably correct and seems to presume a probabilistic or fallibilistic view of knowledge.

Adapted from Kitchener, 1977.

Keeping the major tenets of relativism in mind, that truth exists within a context and may be seen as subjective, King and Kitchener refer to philosophy of science in deriving their upper stages. They observe that life is similar to science in that one must often make decisions on the basis of incomplete or even conflicting information. The scientist is aided in this dilemma by a philosophy of science which helps to provide a way of understanding and judging how one may choose one way of thinking over another. If one cannot say for certain which theory, interpretation, or observation is correct, philosophy of science suggests one might take action on the basis of which is more probable. Appropriately then, King and Kitchener have termed the final stage in Reflective Judgment "probabilism." Probabilistic thinking can be characterized as follows:

1. Judgements are made consciously and deliberated with reflection and thoughtful consideration, are not automatic, imposed, nor unconsidered.
2. Probabilistic judgements are based on reasons and use empirical evidence and/or logic to support conclusions.
3. A person making a probabilistic judgement evaluates evidence from a number of perspectives, acknowledges the relationships among elements, and synthesizes evidence into a coherent, personally meaningful view.
4. Logic and evidence is followed through to a reasoned conclusion by integrating evidence, experts' opinions, and one's own experiences.
5. Reasoning that is probabilistic leads one to form a qualified judgement that, though it may be firmly held,

is not irrevocable and may be revised or reconsidered in the light of new evidence.

In sum, Reflective Judgement is a scheme devised by King (1977) and Kitchener (1977) to trace the development of complex reasoning and judgement-making.

Perry acknowledges a close association between his theory and that of Kohlberg (1969, 1971). Both theories represent stage models. More importantly, however, both describe changes in the way people reason about questions which have no single correct answer. Kohlberg's theory differs, however, in that it deals solely with moral issues. Kohlberg has a six stage theory of moral development. At each stage, there is a different basis for making moral decisions. Moving through the stages in order, those bases are (1) reward, punishment and obedience, (2) instrumental hedonism and concrete reciprocity, (3) maintenance of interpersonal relations and mutuality, (4) maintenance of social order, fixed rules, and authority, (5A) social contract, (5B) higher law and conscience, and (6) universal ethical principles. One hypothesis of this research is that the college years are a time in which persons begin to reason more at the upper stages and less at the lower stages. In other words, it is a time in which people change the way they reason about moral issues. This change in the way they reason about moral issues is linked to changes in the way they make reflective judgments.

Perry also acknowledged links between his theory and Loevinger's theory of ego development. Again, some of the similarity arises from the fact that both are stage theories. However, there are more important similarities. According to Loevinger, cognitive style is one of four ego development strands, and her concept of cognitive style closely parallels Perry's concept of intellectual development. A third link arises from the possibility that the

formation of ego identity is an integral part of Perry's commitment. Table 2 outlines Loevinger's concept of ego development in some detail. In brief, her ego development is a construct

> which assumes that each person has a customary orientation to self and to the world and that there is a continuum (ego development) along which these frames of reference can be arrayed. 'In general, ego development is marked by a more differentiated perception of one's self, of the social world, and of the relations of one's feelings and thoughts to those of others' (Candee, 1974, p. 621) . . . Ego development, then is a central construct which refers to stages of development which include what others have called 'moralization,' 'integration,' 'relatability,' and 'cognitive complexity.' (Hauser, 1976, 929-930).

Empirical Research

As stated above, the research leading up to our project on ego development, moral development, and reflective judgment is too extensive to summarize in any detail. We will focus on longitudinal, cross-sectional, and educational intervention studies, because they seem most relevant to our project.

Various cross-sectional studies of ego development generally have discovered a positive association between age and scores on Loevinger's measure of ego development (Haan, Stroud, & Holstein, 1973; Lambert, 1972; Loevinger & Wessler, 1970; Redmore & Waldman, 1975; Sullivan, McCullough, & Stager, 1970; King, Kitchener, Parker, & Davison, Note 3). All but two of these studies (Haan et al., 1973; and King, et al., Note 3) were carried out over an age range which spanned the junior high and high school years rather than the college years.

In prior work, the principle investigators (King, Kitchener, Parker, & Davison, Note 3) found no differences between high school, college, and graduate students matched in terms of verbal reasoning skills. Table 3 shows the mean ego level scores found by King et al. Haan et al. (1973) found a higher level of ego development in young adults than in late adolescents. Erickson (Note 4) found a gradual rise in ego level in a four year longitudinal study of 20 adult women. The picture that emerges from these longitudinal and cross-sectional studies is that ego level is highly correlated with age (.74 and .64 respectively among boys and girls studied by Loevinger & Wessler, 1970), but that ego level changes slowly with age. For instance, Sullivan et al. found a mean difference of only 1.2 ego stages between youngsters five years apart in age. Erickson (Note 4) found a mean difference of less than a stage in her four year longitudinal study. Redmore and Waldman (1975) found similarly small differences over a 4 year age range (15 to 19 years old).

Given a variable like ego development which seems to change so slowly over time, it seems doubtful that any short term educational program could yield large changes in ego level. Hurt (1974) found that two experimental groups changed significantly and a control group did not. But the three groups did not differ significantly on a post-test, and the largest mean change in any group was less than one-half of a stage. Bereiner (1976) found no change pre- to post- in his study of the effects of education on ego level. Erickson (1973) also found no significant improvement pre- to post- in her experimental group. (She did find that experimental subjects were higher than control subjects on the post-test, but only because control scores dropped pre- to post- whereas experimental subjects remained the same.) From these cross-sectional, longitudinal, and intervention studies, we conclude (a) that ego development changes slowly with age and (b) short-term educational programs have yielded

modest improvements at best in ego level.

The research on moral development during the college years is more extensive. Two of the researchers on this project (Rest, Davison, & Robbins, 1978) have completed a review of longitudinal and cross-sectional studies using the Defining Issues Test, an objective test of moral judgment based on Kohlberg's theory. Even using time intervals as short as two years, consistent upward trends were observed in both cross-sectional and longitudinal studies. Age trends accounted for more variance than either cohort or time of testing effects. Consistent upward age trends have also been observed using Kohlberg's own interview measure (Holstein, 1976; Kohlberg, 1969; Kuhn, 1976), although Kuhn observed no significant change until a full year had passed. Rest (Note 5) found that level of education was more highly correlated with moral judgment scores than was age, a finding which points to an educational influence on moral judgment. In the initial testing of our longitudinal study, we too found significant differences between our high school, college, and graduate students (see Table 3).

A sufficiently large number of studies involving moral education interventions have now been published that Lockwood (Note 6) and Lawrence (Note 7) have completed reviews of these studies. Considering just those studies where the research methodology seemed adequate, Lockwood concluded that such programs seemed to yield small, but significant upward shifts in moral reasoning. Lawrence remained so skeptical about the research methodologies employed, that she drew no conclusion. Even where upward shifts were observed the changes were typically less than one-tenth to one-twentieth of the measure's range. Like Lockwood, we conclude that even short term moral education programs seem to produce small upward shifts in reasoning.

In cross-sectional studies of reflective judgment, Meyer (1975) found significant differences between college freshmen and juniors; the mean

Table 2

Interpersonal Style and Conscious Preoccupations Characteristic of Each Stage in Loevinger's Theory of Ego Development

Stage	Interpersonal Style	Conscious Preoccupations
I-1	Symbiotic	Self vs. non-self
I-2	Receiving, dependent, exploitive	Bodily feelings, especially sexual and aggressive
	Wary, manipulative exploitive	Self-protection, trouble, wishes, things, advantage, control
I-3	Belonging, superficial niceness	Appearance, social acceptability, banal feelings, behavior
I-3/4	Aware of self in relation to group, helping	Adjustment, problems, reasons, opportunities
I-4	Intensive, responsible, mutual, concern for communication	Differentiated feelings, motives for behavior, self-respect, achievements, traits, expression
I-4/5	Add: Dependence as an emotional problem	Add: Development, social problems, differentiation of inner life from outer
I-5	Add: Respect for autonomy, interdependence	Add: Vividly conveyed feelings, integration of physiological and psychological, psychological causation of behavior, role conception, self-fulfillment, self in social context
I-6	Add: Cherishing of individuality	Add: Identity

NOTE: "Add" means in addition to the description applying to the previous level.

NOTE: Adapted from Loevinger (1976) pages 25 and 26.

Table 3

Means and Standard Deviations of the Concept Mastery Test, Defining Issues Test of Moral Development, Sentence Completion Test of Ego Development, and Reflective Judgment Interview in the Initial Testing of Study 1

	Concept Mastery Scores by Sample		
	High School	College	Graduate
$\bar{X}$**	42.26	84.11	111.37
S	16.83	23.36	33.34

	Defining Issues Test Scores by Sample		
	High School	College	Graduate
$\bar{X}_D$** SCORE	23.39	26.61	30.40
S_D SCORE	5.38	6.70	6.18

	Sentence Completion Test Scores by Sample		
	High School	College	Graduate
$\bar{X}$	35.23	37.94	36.05
S	4.85	4.35	4.27

	Reflective Judgment Interview Scores by Group		
	High School	College	Graduate
$\bar{X}$**	2.77	3.64	5.67
S	.49	.81	.92

**Group means were significantly different at the .01 level.

difference corresponded to about 1 stage position. Even after controlling for differences in verbal reasoning abilities, Kitchener (1977) found significant differences between high school juniors, college juniors, and graduate students (see Table 3). Blake (Note 9) found a small, but significant upward trend in average Perry position scores from the freshman to the junior years with a slight drop during the senior year. Kurfiss (1975) found no differences between her freshmen and juniors, but her measuring instrument was quite unlike that used in other studies. With the exception of the Kurfiss study, older, better educated subjects seem to display higher stage levels of reflective judgment. Mean differences seldom exceed one stage over a two year period, except in Perry's original study. It is not clear to what extent these age-education differences are due to maturation or education.

A number of educational interventions have been based on Perry's scheme (Widick, Knefelkamp, & Parker, 1975; Touchton, Werthermer, Cornfield, & Harrisen, 1976; Stephensen & Hunt, 1977; Widick & Simpson, Note 10). Where control groups were used, the experimental groups seemed to have shifted upward more than controls. But the significance of these differences was not always tested. None of the mean differences were as large as a single stage. Results of these studies weakly suggest the conclusion that short-term interventions have small upward effects on reflective judgment styles.

Measurement Methods

We used the Reflective Judgment Interview and three questionnaires, the Concept Mastery Test, the Defining Issues Test, and the Sentence Completion Test of Ego development. All four instruments will be described in alphabetical order. But first a word about why we have selected the three measures of cognitive-social development which we have chosen. Questions having no single answer arise in many (not necessarily mutually exclusive) domains. The three

cognitive-social measures were chosen to assess reasoning in several of these domains. The Sentence Completion Test of ego development deals primarily with personal issues. The Defining Issues Test deals with moral issues, and the Reflective Judgment Interview deals with social, historical, religious and scientific issues. We chose these cognitive-social measures because they seemed to tap complex reasoning about complex intellectual problems in a broad range of domains.

The Concept Mastery Test (Terman, 1973) is an objective, paper-pencil test of verbal reasoning ability, the ability to solve verbal puzzles which have a single correct answer. The kinds of questions included on the test closely parallel those found on standard paper-pencil measures of verbal intelligence, except that the Concept Mastery Test (CMT) is designed to discriminate among people at the very highest levels of verbal reasoning. It contains two sections. In Section I, the examinee is presented two words and must correctly determine whether those two words are synonyms or antonyms. In Section II, the examinee must correctly answer analogous puzzles, such as "cat is to kitten as dog is to (puppy)." While the reliability varies depending on the sample and the type of coefficient used, the internal consistency reliabilities generally seem to be in the low .90s while the test-retest reliabilities generally fall in the .80s (Terman, 1973). The test takes about 40 minutes to administer. Where group differences were found on other measures, the analysis of covariance and CMT scores were used to assess whether or not those group differences on other measures could be attributed to differences in conventional verbal reasoning skills. CMT scores were also used to assess whether people entering different educational programs begin with different levels of verbal reasoning skill.

The Defining Issues Test (DIT) is an objective test of moral development based on Kohlberg's theory (Rest, Cooper, Coder, Masanz, & Anderson, 1974;

Rest, 1975; Rest, Note 11). It is composed of six stories each describing a moral dilemma. Following a dilemma are 12 moral issues each corresponding to a stage in Kohlberg's theory. The examinee must rate each issue in terms of importance and rank order his/her four most important issues. The test yields 6 stage scores, each indicating how much importance the subject attaches to issues keyed to Stages 2, 3, 4, 5A, 5B, and 6 respectively in Kohlberg's theory. In addition it yields a P score which is simply the sum of stage scores 5A, 5B, and 6 and which indicates how much importance the subject attaches to Kohlberg's Principled Issues (issues characteristic of Stages 5 and 6). And it yields an index of overall moral development level, called the D score. Davison and Robbins (1978) report that the internal consistency reliability of the D score is in the .80s while test-retest reliabilities generally fall in the .70s or .80s. The test takes 40 to 50 minutes to administer.

The Reflective Judgment Interview was developed by two students of the principle investigators, King (1977) and Kitchener (1977). It consists of four intellectual problems with no unique correct answer; one problem in historical interpretation, one on the objectivity of journalism, one on scientific and religious explanations of creation; and one on chemical additives in foods. Each problem contains two statements describing opposing positions on the issue. After hearing each problem, the examinee verbally answers several questions, such as "What do you think about these statements?", "On what do you base your point of view?", and "Can you ever know for sure that your position is correct?". Responses to each problem are scored for the position in the scheme which best seems to characterize the response. An examinee's total score is the average of the

four problem scores. King (1977) obtained internal consistency (Cronbach's alpha) reliability estimates ranging from .84 to .96 depending on the group. No test-retest reliability estimate is available. Interrater reliabilities ranged from .68 to .96 and interrater agreement coefficients (Lawles & Lu, 1972) ranged from .59 to .77.

The last instrument to be used in this research is Loevinger's Sentence Completion Test of Ego Development (Loevinger & Wessler, 1970; Loevinger, Wessler, & Redmore, 1970). There are two forms of this instrument, one for men and one for women. Both forms contain 36 sentence stems to be completed by the examinee. Each sentence completion is rated for ego level (a level in Loevinger's theory) and the 36 item scores are combined into a composite. Loevinger & Wessler (1970) report an interrater reliability of .86 and an interrater agreement of 61%. Redmore and Waldman (1975) obtained internal consistency reliabilities ranging from .80 to .89. Test-retest reliabilities in this same study ranged from .44 to .91, depending on the sample and the scoring method used.

This report focuses on three major studies in our project. The first is an extension of a longitudinal study in which James Rest examined the moral reasoning development of people who did and did not go on to college after high school to assess whether one or the other group was changing faster. The study covered the four years immediately following high school. The study is more fully reported in Project Report #1, "The impact of higher education on moral reasoning development."

The second study summarized here is more fully reported in Project Report #3, "Reflective judgment and its relationship to academic field." It is a cross-sectional comparison of students in engineering majors and liberal arts majors on two instruments, the Concept Mastery Test and

the Reflective Judgment Interview. The purpose of this study was to assess whether freshmen to senior changes on the two instruments were larger in the liberal arts majors than in the engineering majors. The last study is a cross-sectional comparison of people with bachelor's degrees who did and did not go on to graduate school. All four measures, the Concept Mastery Test, the Reflective Judgment Interview, the Defining Issues Test, and the Sentence Completion Test were used in this study. Our purpose was to investigate whether those who went on to graduate school seemed to be undergoing larger changes.

The remainder of this report is divided into five sections. The first three describe the three studies outlined above. The fourth summarizes our attempts to disseminate project results. And the last gives our conclusions and suggestions for future research.

Study I

In his study of children's moral judgment, Piaget (1932/1965) emphasized peer interaction as the crucial experience leading to development. Kohlberg (1969) hypothesized that not only peer interaction but also "role taking" experiences contributed to moral judgment development. "Role taking" experiences include taking responsibility for others, participation in social organizations and institutions, democratic parent-child discussions, as well as give-and-take among peers. Neither Piaget nor Kohlberg emphasizes higher education as a facilitator of moral judgment development. Their deemphasis on formal education possibly derives from the emphasis that the "socialization" view placed on direct teaching--cognitive developmentalists want to depict development in terms of internal transformations and reconstructions of experience, not in terms of direct learning.

In moral judgment research using the Defining Issues Test ("DIT"),

higher education has been a strong correlate of moral judgment. Table 4 shows the average DIT scores for four student groups: the more educationally advanced groups have higher moral judgment scores.

Table 4

Average Moral Judgment Scores of Student Groups

Student Group	Average DIT (P-index)	Number of Subjects in Sample
Junior high students	21.9	1322
Senior high students	31.8	581
College students	42.8	2479
Graduate school students	53.3	183

In several studies of adults, the correlations of DIT scores with education are higher than for age: Coder (1975) found a negative correlation with age (r = .10) and a positive correlation with education (r = .25); Crowder (1976) found correlations of -.05 and .25, respectively: G. Rest (1977) found a correlation of .45 with education; and Dortzbach (1975) found a negative correlation with age and a positive correlation with education. Furthermore, cross sectional data suggest that adults seem to plateau in moral judgment development after they end their formal schooling. If adults are grouped according to the highest level of formal education attained, the average DIT scores are comparable to the average DIT scores of students currently at that level (for instance, current high school

students average about the same as adults with high school educations but who are in their 40s, 50s, or 60s). In other words, formal education seems to be strongly associated with moral judgment development in these cross-sectional studies.

The purpose of the present study was to more adequately explore the relationship between formal education and moral judgment development by using longitudinal data. The major questions were: after high school graduation, do subject who go to college show greater development in moral judgment than subjects who do not go to college? To what influences do subjects themselves attribute their own change in moral judgment?

Method

59 subjects in an ongoing longitudinal study have been tested since 1972 at two year intervals. The available data for this report are three testings: as seniors in high school, two years after graduation, and four years after high school graduation. 38 subjects had gone on to college and 18 had not. Differences in the course of development of these two groups are the major interest of the study. Subjects have completed each time a questionnaire package including the DIT, Comprehension of Moral Concepts, and a political attitudes test ("law and order" test). Also, subjects completed a short "life history" questionnaire, and were asked to describe what experiences or influences had affected their moral thinking (see Rest, 1979 for details).

The DIT is a multiple-choice test of moral judgment development derived from Kohlberg's approach. Subjects rate and rank stage-prototypic items according to their perceived importance in making a decision about a hypothetical moral dilemma. Developmental indices are based on the way subjects

differentially rate and rank the items. The major indices are the P score (relative importance given to "Principal Moral" items, Stages 5 and 6), and the D score (a composite score based on scaled values of items).

Results and Discussions

Both the P scores and D scores showed similar patterns of development distinguishing the college from the noncollege groups. In high school (when all subjects had the same level of formal education) the college-bound group was not significantly different from the non-college-bound group. Then two years after high school graduation, both groups had gained significantly on moral judgment and were still not significantly different from each other, but by the time that four years had passed after high school, the two groups were significantly different and were showing divergent patterns of development. Figures 1 and 2 show the patterns of change for the P and D scores over time. Tables 5 and 6 show the averages for each group at each testing, simple t-tests at each testing between the college and noncollege groups, and ANACOVA results which contrast the college and noncollege groups after controlling for the pretest scores (scores obtained in high school).

The college subjects also showed a divergent pattern of development from the noncollege subjects on political attitudes, but not on the Comprehension test. (Problems in reliability on the Comprehension test may account for its nonsignificant differences between the groups, although Comprehension scores did significantly increase for both groups over the four years.)

From the "life History" data, it was seen that some subjects were still living with their parents (n = 24) whereas others were not (n = 34). Comparison of the developmental patterns of these two groups showed that those subjects living away from home started out with a developmental advantage, maintained

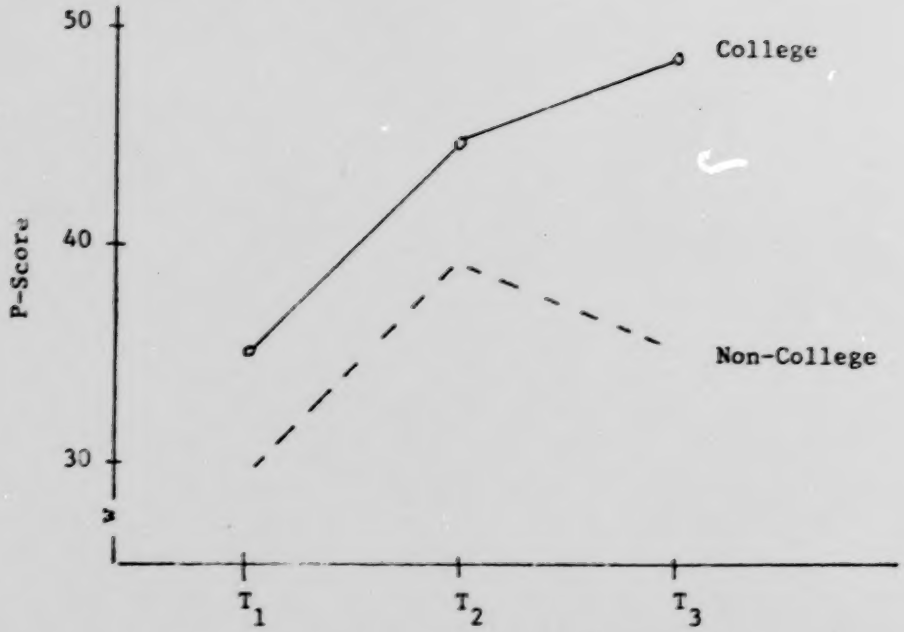

Figure 1 Comparison on P-Score of College and Non-College Groups

Table 5

Comparison on P score of College and Non-College Groups

	Means T_1	Means T_2	Means T_3
College (n = 38)	34.8	44.4	48.5
Non-College (n = 18)	29.3	38.9	35.1
t-tests	t = 1.43	t = 1.67	t = 3.37
Coll. v. Non-Coll.	p = .16	p = .10	p = .002
ANACOVA	_____	F = .90	F = 9.2
Coll. v. Non-Coll.	_____	p = .35	p = .004
(partialling out T_1 scores)			

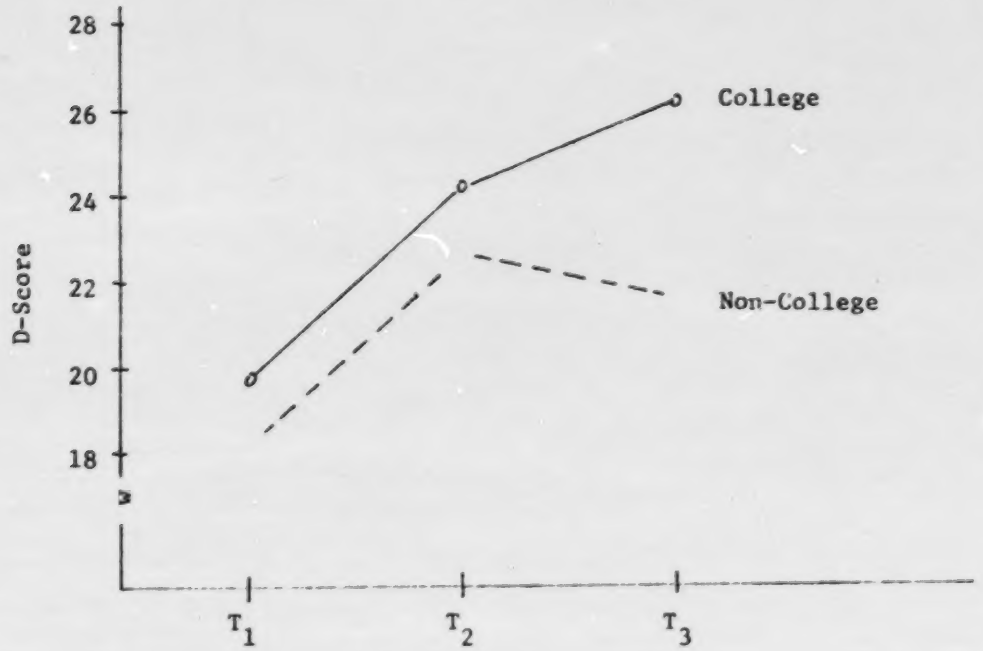

Figure 2. Comparison on D-Score of College and Non-College Groups

Table 6

Comparison on D Score of College and Non-College Groups

	Means T_1	Means T_2	Means T_3
College (n = 38)	19.8	24.3	26.3
Non-College (n = 18)	18.6	22.9	21.3
t-tests	*t* = .83	*t* = .76	*t* = 2.35
Coll. v. Non-Coll.	*p* = .41	*p* = .45	*p* = .02
ANACOVA	_____	F = .26	F = 9.95
Coll. v. Non-Coll.	_____	*p* = .62	*p* = .001
(partialling out T_1 scores)			

that advantage over the subsequent four years, and increased more than those subjects living at home. Analysis of the interaction of being in college (or not) and living at home (or not) showed that each condition had major effects and they were additive (the greatest development showing up in college students not living at home, the least in noncollege subjects living at home, with the other two combinations in between). (See Tables 7 and 8)

College-bound subjects were subdivided according to the type of institution attended: either a large research oriented university (n = 17) or a small, four-year college (n = 21). No difference in developmental patterns was found.

Table 7

Comparison on P Score of Subjects Living at Home Versus Those Not Living at Home

	Time 1	Time 2	Time 3
Living at Home (n = 24)	31.9	38.5	39.8
Not Living at Home n = 34	34.3	45.4	48.5

Table 8

Comparison on P Score of Subjects Divided Into Place of Residence and Education

Group	Time 1	Time 2	Time 3
In College, Living at Home (n = 14)	31.3	39.5	45.6
In College, Not at Home (n = 24)	37.0	46.2	49.7
Not in College, at Home (n = 10)	30.7	38.8	34.4
Not in College, Not at Home (n = 8)	27.5	39.0	36.1

Subjects were asked to describe the influences on their moral thinking. What is it that subjects perceive in their environments and experiences that seems to them to cause change? The responses were classified into 18 categories, including reading, formal instruction, current events, new social

contacts, marriage or job responsibilities, making decisions "on my own," living away from parents, religious experiences or instruction, direct involvement in political or community affairs, personal tragedy, travel, etc. (See Table 9). The most frequently mentioned influence was "new social contacts" (n = 28), however the developmental progress of these subjects actually lagged behind the subjects not mentioning this influence but mentioning other influences. 22 subjects attributed influence to "new real world responsibilities" (marriage, job, managing money, children), however the development of this group did not end up greater than subjects not mentioning this influence. Fewer subjects cited the influence of each of the other factors, but mentioning some of these other factors was associated with more dramatic growth patterns: subjects who attributed changes in their thinking to spending more time contemplating moral issues did in fact show more dramatic development than subjects not citing this factor. Similarly, more dramatic growth patterns were associated with subjects citing direct involvement in community/political affairs, living away from parents, and travel in contrast to subjects not citing those influences.

In summary, this study supports the notion that higher education fosters development in moral judgment, particularly at the four year mark rather than in the first two years of undergraduate education. In addition to this project six other longitudinal studies of moral judgment using the DIT have been completed on college subjects and all report significant upward trends (although these studies do not contrast the college students to noncollege students): Broadhurst (1980); Kaseman (1980); Mentkowski (in press); Sheehan, Hustad, and Candee (1981), Whiteley (in press). The vexing problem remains in

clarifying how higher education has its effects. It may very well be the case that different aspects of the experience have different impact and salience for different subjects. Such a point of view is common sensical, however our present research strategies are designed differently: to find common effects in groups of subjects from the same causes. Progress in this research is likely to come from developing a research methodology which enables us to reliably draw inferences from the intensive, idiographic examination of individual subjects, and which collects data that are not so indirect and not so filtered through the subject's own retrospective theories about what affected him or her.

Table 9

Comparison on P Scores of Subjects Who Attribute Change to Various Life Experiences

Type of Life Experience	Number of Subjects Citing This Influence		Time 1	Time 2	Time 3
1. Reading	13	"YES" Group	29.2	38.9	41.3
		"NO" Group	34.4	43.5	47.3
2. Formal instruction or study	16	YES	29.1	41.4	44.5
		NO	34.9	42.8	46.6
3. Current issues and events	19	YES	35.1	42.9	46.5
		NO	32.4	42.3	45.7
4. Spending more time contemplating issues	8	YES	34.6	48.8	53.6
		NO	33.1	41.5	44.8
5. Maturation, "getting older," sense of "growing up"	5	YES	24.3	34.0	29.0
		NO	34.1	43.2	47.6
6. New social contacts, an expanding social world	28	YES	31.9	41.7	39.4
		NO	28.8	43.1	51.9
7. Specific influential people	3	YES	37.8	42.2	41.7
		NO	33.0	42.5	46.2
8. New "real world" responsibilities--marriage, job, managing money, children	22	YES	38.3	38.5	45.5
		NO	30.3	44.8	46.3
9. "Making decisions on my own"	11	YES	42.1	45.8	45.5
			31.3	41.7	46.1
10. Making decisions for the future	3	YES	35.0	34.5	40.0
		NO	33.2	42.9	46.3
11. Living away from home	11	YES	42.0	50.0	52.1
		NO	31.3	40.7	44.6
12. Religious experiences and/or instruction	4	YES	16.7	35.0	40.0
		NO	34.5	43.0	46.4
13. Direct involvement in community/world political affairs	3	YES	31.1	44.5	55.0
		NO	33.3	42.4	45.5
14. Experiencing or witnessing personal tragedy	5	YES	35.3	45.7	39.0
		NO	33.1	42.2	46.6
15. Particular time of personal stress as a turning point	1	YES	23.3	35.0	46.7
		NO	33.5	42.6	46.0
16. "Change in lifestyle"	2	YES	38.3	39.2	47.5
		NO	33.1	42.6	45.9
17. Travel	3	YES	31.1	53.9	58.9
		NO	33.4	41.9	45.3
18. No change in thinking	6	No change	31.7	41.4	45.7
		Other	33.5	42.6	46.0

Study II

Nearly every college catalogue names the development of the capacity to form thoughtful judgments about the complex problems of modern society as a central mission of higher education. Unfortunately, few institutions have attempted systematic research to ascertain their success in reaching this goal. Educators have most frequently examined gains in achievement and verbal reasoning test scores and changes in scores on traits like autonomy, complexity, social liberalism and interest in intellectual and cultural activities to find evidence of an increased capacity to form thoughtful judgments. However, although each of these measures is clearly related to the kind of intellectual changes to which the catalogues refer, none directly assesses the college student's ability to think about complex issues. For example, the items on verbal reasoning tests have single, easily verifiable right and wrong answers; the complex problems of modern society do not have such undeniably compelling solutions. In an era in which the demands upon higher education to account for its outcomes are steadily increasing, the need to more clearly document intellectual development during the college years becomes critical. The recent work of King (1977) and Kitchener (1977) derived from William Perry's (1970) research at Harvard provides a theoretical model and an instrument that appears promising in measuring the impact of higher education on intellectual development. The present study examined the relationship between year in college, academic major, academic performance, and scores on King and Kitchener's measure of intellectual development. In general, although it is premature to make definitive statements about the long range contribution of these models, the initial results offer encouragement to educators and psychologists alike.

Statement of the Problem

Too little systematic attention has been paid to the elements in the educational environments associated with intellectual development. The results of the King and Kitchener studies for the college sample demonstrate the need to identify those elements. The scores of the college students on the Reflective Judgment Interview were more variable than the high school or graduate samples and less closely associated with verbal reasoning ability than the other groups. Sex differences also approached significance. In addition, their scores on the RJI were more similar to the high school students than the graduate students. In an effort to make sense of the scores of the college juniors, Kitchener (1977) suggested that the variability over the sample could be related to the major fields of the students. She observed a trend in her findings for students in scientific and technical majors to score lower than those with liberal arts majors. Other research on intellectual development, using a different instrument (Pohl and Pervin, 1968) found a significant association between major and cognitive complexity and made Kitchener's hypothesis plausible. It was the purpose of this study to systematically test the relationship between Reflective Judgment and academic major and to explore the relationship between Reflective Judgment, grade point average, and satisfaction with choice of major in order to begin to delineate the factors in the college environment related to level of intellectual development. This study also replicated the findings of Strange (1978) of a relationship between year in college and score on the RJI, and explored further the relationship between verbal reasoning ability and Reflective Judgment.

Method

Subjects

The sample in this cross-sectional study consisted of 64 students randomly selected from computer listings of University of Minnesota students enrolled in the fall of 1978. The subjects were equally divided among four groups: freshmen with engineering majors, seniors with engineering majors, freshmen with humanities/social sciences majors, seniors with humanities/social sciences majors. Each group contained equal numbers of males and females. All students had attended a Minnesota high school and all seniors had declared their majors at least three quarters prior to fall, 1978.

Instruments

Three instruments were administered to each subject, the Reflective Judgment Interview, Terman's Concept Mastery Test, and a General Information Sheet. Preliminary Scholastic Aptitude Test scores were obtained from admissions data.

Reflective Judgment Interview

The Reflective Judgment Interview includes four dilemmas presented in random order to an individual subject and a series of probe questions that follow each dilemma. The measure takes approximately 45 minutes to complete and is administered by a trained interviewer. The format is considered semi-structured since the interview varies from the standard probe questions when necessary to clarify a subject's response. The interview is taped and a transcription is rated by trained judges according to the scoring rules developed jointly by King and Kitchener. All identifying information is edited from the transcript to avoid rater bias. Each of the dilemmas involves a controversial issue of general interest. The dilemma presents two points of view on the topic and the probe questions solicit the person's

perspective on the issue and rationale for the conclusion reached.

The four dilemma format was developed by King and Kitchener after a pilot study of a six dilemma format. Results from the pilot were also used in establishing the scoring rules.

Although the instrument is too recently developed to have accumulated a large body of reliability and validity evidence, the initial results are supportive in two ways. First, the coefficients obtained are quite high. Second, the information is encouraging across a range of measures of reliability and internal consistency. For the King and Kitchener sample, inter-rater reliability, item-total correlations, inter-rater agreement calculated in terms of significant differences between rater, and percent of agreement between raters all suggested high correspondence between the two judges' ratings. Specifically, the inter-rater reliability coefficients ranged from .68 to .94, item total correlations from .90 to .92 and overall inter-rater agreement from .75 to .85. Overall differences between judges were not significant. They also used internal consistency in subjects' scores across dilemmas as an indicator of construct validity, arguing that high correlations between scores on the four dilemmas for each subject would support their claim for validity of the theory. For the total group, correlations ranged between .84 and .88 and within groups, between .56 and .86. All of these correlations were significant ($p < .01$). For the College of Agriculture sample, inter-rater reliability was .63 ($p < .001$) with 77% of the ratings falling within the 2/3 point criterion level for inter-rater agreement.

Concept Mastery Test

The Concept Mastery Test, first developed by Terman in 1947, is a 40 minute written test of vocabulary and verbal intelligence. Because high scores are dependent on a large vocabulary as well as the ability to reason with

vocabulary, it is appropriately referred to as a measure of verbal reasoning. The test is divided into two sections: synonyms-antonyms and analogies. The test manual (1973) suggests that it is suitable for use with college students, graduate students, and gifted high school students. Terman reports correlations of between .55 and .69 with other measures of verbal abilities, and somewhat lesser correlations with I.Q. and numerical abilities measures. Kitchener (1977) reported a non-significant correlation of .23 with the Reflective Judgment Interview for her liberal arts college junior sample, but much higher correlations for the high school (.63) and graduate student (.78) samples. The test is designed for either group or individual administration; in this study, the test was individually administered. The CMT in particular was selected because of its appropriateness for a college population likely to have high levels of verbal reasoning ability and also because using the same instrument as Kitchener (1978) allowed for easier comparability of findings.

General Information Sheet

The General Information Sheet was used to obtain demographic information, academic achievement information, and data on choice of major field. The instrument contains three pages. The first page asks for name, address, and telephone number. The second page includes items asking for date of birth, residence, college, major, year in school, and parents' education and occupations, based on Hollingshead's (1965) two factor index of social position. Subjects were requested to place an "X" in the blank that best describes their parents' education and occupations at the time they entered college. The final page contains items related to questions relating to grade point average and choice of major field. All subjects were asked to

describe in a few sentences their rationale for their choice of major or level of satisfaction with their choice of major. The questionnaire took approximately ten minutes to complete.

Procedure

Testing was conducted in one two-hour session for each subject. Each session began with a reading and signing of the informed consent in which the nature of the research was briefly described. Half of the subjects received the Concept Mastery Test first, half the Reflective Judgment Interview first, and all completed the General Information Sheet as the final task. The Reflective Judgment dilemmas were presented in random order.

All testing took place in a soundproof room at the University of Minnesota, equipped with a desk, two chairs, a cassette tape recorder and tapes, and the testing materials. A separate cassette tape was used for each dilemma. Each tape was identified by dilemma number and by the subject's code number which was known only by the investigator. The data collection was completed over a six week period between November 1, 1978 and December 10, 1978.

All individuals were instructed that the CMT was a general test of verbal reasoning and the RJI, a measure of how they think about issues of general concern.

Data Analysis

Both analysis of variance and covariance were used to analyze the data. The independent variables were major field and year in college. Concept Mastery Test score and Preliminary Scholastic Aptitute Test score were the covariates, and Reflective Judgment Interview score the dependent variable. Both an analysis of variance and covariance were chosen because in the Kitchener (1977) study CMT score was shown to be significantly related to RJI

score. Therefore, without controlling for verbal reasoning ability, any significant results could be interpreted as the result of differences in verbal reasoning as much as Reflective Judgment. Similarly, since some earlier research reported a moderate relationship between intellectual ability and measures of ego, moral, and cognitive development (Loevinger & Wessler, 1970; Kohlberg, 1969; Rest, 1976; Schroder, Driver, & Struefert, 1967) and since Kitchener also found a significant association between intellectual ability and RJI score, PSAT served as a second covariate in this study. Pearson product moment correlations were used to test the relationships between GPA in major, satisfaction with choice of major, and Reflective Judgment level.

Additionally, since sex differences on overall RJI score reached significance for Strange (1978) and approached significance ($p < .07$) for the Kitchener sample, a post-hoc analysis of sex differences on RJI scores was done. The possibility that the sex differences may be better explained through reference to academic major (Kitchener, 1977) and the lack of similar findings in the Perry research were the rationale for the decision in favor of a post-hoc analysis rather than an hypothesis test.

Results

Three independent judges were involved in the scoring of the RJI. One judge rated all 64 interviews (256 dilemmas) and the other two each rated half (128 dilemmas). Before the transcripts were sent to the raters, they were read and edited for cues that would be indicative of the subject's age, sex, educational level or major field.

The same rating procedure that was used in prior studies was employed here. Each dilemma was assigned three scores. As explained by King in reference to her study,

> Three scores were used to give the most accurate representation possible to characterize the Reflective Judgment position at which the subject discusses each dilemma. The triple score helps identify and describe the variance across positions and allows a rater to make a more complex assessment of a subject's reasoning style. (p. 121)

If the average of the three scores assigned to any dilemma by the judges differed by more than one point, the judges were asked to re-rate those dilemmas. Forty-four of the 256 dilemmas required a second rating. If, after a second independent rating, the discrepancy was not resolved, those judges discussed that dilemma and came to an agreement. Two dilemmas required this second re-rating.

Reliability of the Reflective Judgment Interview (RJI)

The overall reliability coefficient between Judge 1 and Judge 2 was .84 and between Judge 1 and Judge 3, .89. Since judges 2 and 3 rated no common protocols, their inter-rater reliability could not be computed. Inter-dilemma reliability coefficients ranged from .17 to .38. The inter-rater reliability coefficients are comparable to those reported in earlier studies but the inter-dilemma correlations are lower than those obtained by Kitchener (1977), whose coefficients ranged from .56 to .86 for the college sample, and also somewhat lower than those reported by Strange (1978), .38 to .63. This difference may be partially explained by the smaller range in scores over the total sample in this study, since a reduced range in scores depresses the correlations obtained. These lower inter-dilemma correlations may also be the result of the uneven nature of intellectual development among college students, or may be a consequence of a weakness in interviewer training,

interviewer bias, or rating skill. Dilemma-total correlations, in which a judge's mean score for a dilemma was correlated with the subject's total score, were moderate, ranging from .35 to .52.

Inter-rater agreement was calculated in two ways. First, simple percent of agreement between raters (in their original ratings) was computed, with agreement defined as one point or less discrepancy between the averages of the judges' ratings for each dilemma. Then, using the procedure recommended by Lawlis and Lu (1972), these coefficients were corrected for chance agreement. Both sets of coefficients are presented in Table 10. When corrected for chance agreement, the percent of agreement is best described as moderately high. In sum, these data indicate that the RJI reliable assessed subjects' responses.

Table 10

Reflective Judgment Interview Inter-Rater Agreement[a] Coefficients for Total Sample: Unadjusted and Adjusted for Change Agreement

Dilemma	Unadjusted	Adjusted
1 (Egyptian Pyramids)	.813	.762
2 (News Reporting)	.782	.753
3 (Creation/Evolution)	.844	.711
4 (Chemicals in Foods)	.767	.692
Overall	.802	.728

[a]Agreement defined as one point or less discrepancy.

The distribution of mean RJI scores is presented in Table 11. The range of mean scores for the total sample is from 3.13 to 5.08, with the grand mean at 3.81. The scores of the majority of freshmen fall into the dualistic category while the majority of seniors score at the relativistic level. The freshmen in humanities and social science majors had the lowest mean score (3.52) while the seniors in humanities and social sciences had the highest mean score (4.05). In general, these scores are comparable to those reported in the other research on Reflective Judgment. Kitchener (1977) obtained a mean of 3.65 for her sample of 20 college juniors in liberal arts and King and Parker (1978) reported a mean of 3.93 for their sample of 20 juniors in agriculture. The range of scores here is slightly smaller than the range reported by Strange (1978) in his sample of traditional age freshmen and seniors at the University of Iowa, 3.28 to 5.56.

Group Differences in Reflective Judgment Interview Scores

One major purpose of this study was to examine the relationship between the level of intellectual development of individuals in college and their year in college and academic major. Prior research (Strange, 1978) had indicated significant differences between freshmen and seniors and had pointed to a possible relationship between academic major and level of development (Kitchener, 1977). In this section, the results most relevant to those issues will be discussed. Hypothesis I dealt with the differences between freshmen groups in RJI score. A significant relationship between RJI score and major for these first quarter freshmen would suggest an association between Reflective Judgment level and initial choice of major. An analysis of variance indicated no significant differences between the two freshmen groups on the RJI (Table 12). No evidence of influence of Reflective Judgment level on choice of major at entrance to college has been revealed in this study.

Table 11

Distribution of Mean Reflective Judgment Interview Scores

Over All Dilemmas for the Four Student Groups

(by percent)

Mean Reflective Judgment Scores		Student Groups: Humanities/ Social Science Freshmen	Engineering Freshmen	Humanities/ Social Science Seniors	Engineering Seniors
	1.5				
	2.0				
Dualism	2.5				
	3.0	44	6	6	6
	3.5	38	56	25	44
	4.0	19	38	31	25
	4.5			25	13
Relativism	5.0			13	13
	5.5				
	6.0				
Probabilism	6.5				
	7.0				
		100%	100%	100%	100%
		$\underline{N}$ = 16	$\underline{N}$ = 16	$\underline{N}$ = 16	$\underline{N}$ = 16
		$\bar{X}$ = 3.52	$\bar{X}$ = 3.72	$\bar{X}$ = 4.05	$\bar{X}$ = 3.94
		S.D. = .329	.317	.578	.557
	Range =	3.15-4.17	3.21-4.50	3.25-5.04	3.17-5.08

Note: Totals may not equal 100 due to rounding.

Table 12

Analysis of Variance on Reflective Judgment Interview Scores for Freshmen by Group and Sex (N = 32)

Source of Variation	df	ms	F	P
Group	1	.313	2.853	.102
Sex	1	.014	.126	.725
Group x Sex	1	.049	.445	.510
Residual	28	.110		
Total	31			

Hypothesis II focused on the difference between the seniors in engineering and the seniors in humanities and social sciences. Any significant difference between the two groups on RJI score would suggest that academic major is related to the level of intellectual attainment. The results of a one way analysis of variance presented in Table 13 show no significant differences between these two groups. Although the mean score of seniors in humanities/social sciences is higher than their counterparts in engineering, the difference is not significant. Nothing in these findings supports the interpretation that the academic majors produce different levels of intellectual development as measured by the RJI.

The findings for Hypothesis III, however, lead to rejection of the null hypothesis. When the scores of freshmen and seniors were compared, a statistically significant difference resulted. A three-way analysis of variance with year in college, major, and sex as the independent variables and RJI score

Table 13

Analysis of Variance on Reflective Judgment Interview Scores for Seniors by Group and Sex (N = 32)

Source of Variation	df	ms	F	P
Group	1	.096	.294	.592
Sex	1	.521	1.598	.217
Group x Sex	1	.018	.054	.818
Residual	28	.326		
Total	31			

as the dependent variable revealed a significant main effect for year in college at the $p < .01$ level, but no significant interaction effects (Table 14). In other words, the differences between freshmen and seniors in humanities/social sciences are not significantly greater than the differences between freshmen and seniors in engineering. The outcome of this analysis supports the findings of Strange (1978) that a sample of older, more educated students score higher on the RJI than younger and less educated students.

Effects of Controlling for Verbal and Intellectual Skills

Since both PSAT and CMT were significantly associated with Reflective Judgment level in Kitchener (1977) and ACT composite score correlated significantly with Reflective Judgment in Strange (1978), the relationships between RJI, PSAT, and CMT were tested in this study. For this sample no

Table 14

Analysis of Variance on Mean Reflective Judgment Interview Scores by Year in College, Major, and Sex

Source of Variation	df	ms	F	P
Year in College	1	2.250	10.324	.002
Major	1	.031	.144	.706
Sex	1	.353	1.618	.209
Year in College x Major	1	.378	1.733	.193
Year in College x Sex	1	.182	.837	.364
Major x Sex	1	.004	.018	.894
Year in College x Major x Sex	1	.062	.287	.594
Residual	56			
Total	63			

significant correlations between RJI, CMT, and PSAT were found with the exception of the correlation between the two ability measures, the PSAT and CMT. When these results are compared with those obtained by Kitchener for the college sample only, they appear less discrepant. In her research the correlation between RJI and CMT was .23 and between RJI and PSAT, .30. Neither correlation reached statistical significance. Both the correlations obtained in this study and the results of the analysis of covariance suggest that the Reflective Judgment Interview is measuring a construct distinct from

verbal reasoning ability. The differences between groups remained significant after CMT and PSAT are covaried out. These findings support Kitchener's interpretation that Reflective Judgment and verbal reasoning skills develop at different rates during the college years.

The mean score of the four groups on the CMT were as follows: engineering freshmen, 50.2; engineering seniors, 55.6; humanities/social sciences freshmen, 54.5; and humanities/social sciences seniors, 64.1. No significant differences on the CMT overall were obtained, but when the scores on each half of the CMT were analyzed separately by a one way analysis of variance, differences between engineering and humanities/social sciences majors were revealed. The students in liberal arts majors scored significantly higher on the synonyms-antonyms section ($p < .05$) than the engineering students, a section of the test that Terman asserts is a measure of vocabulary predominantly. This difference is congruent with the increased emphasis in liberal arts on reading and writing skills. No significant differences between groups on PSAT scores were found even though the mean score of those in engineering was slightly higher (51.6) than the mean of those in humanities/social sciences (50.4).

Discussion of the Results

The main effect for year in college supports the findings of other RJI researchers and more generally supports the trend in the literature on college student development which demonstrates an impact of college on those who attend and persist beyond what could be expected by maturation alone. However, the range in scores and the highest scores are smaller and lower than what would be predicted from prior research on RJI and considerably smaller than the range in the original Perry sample.

A recently published study by Astin (1978) appears useful in making sense of the small range in scores. Astin's work also seems helpful in interpreting the relationship between Reflective Judgment, year in college, and academic major. Astin and his colleagues studied 200,000 individuals longitudinally between 1961 and 1974, both non-students and students in a variety of higher educational institutions. He attempted to ascertain the relative impact of characteristics, and the amount and type of involvement in college life was measured using type of residence (dormitory, apartment, or at home with parents), participation in special academic programs and research activities, faculty-student interaction, participation in social and athletic activities and sororities and fraternities, and academic involvement (time studying) as criteria for involvement. He reported that the size of the institution was correlated with its impact on students. Large public universities with a high ratio of commuters and fewer opportunities for involvement in special programs or campus activities, or for faculty-student interaction showed less change on both cognitive and affective dimensions. The University of Minnesota is typical of the large universities he studied. Therefore, it is reasonable to suggest that the nature of the institution from which the subjects in this study were sampled may have depressed the scores on the RJI, and by depressing the scores of the seniors, reduced the range of scores in the total sample.

Furthermore, when Astin analyzed the influence of major field he found those in engineering showed smaller than average changes and those in social sciences, greater than average changes. Engineering students, he reported, spent proportionally more time studying, had more stringent grading requirements, and were frequently isolated from students in a separate

technological school. As a consequence of the difficulty of the academic program and the separation, engineering students were less involved in college life and in turn, showed less change. The small differences between RJI scores of the two engineering groups, less than one quarter position, is consistent with Astin's findings. Since no data are available on engineering majors from a smaller college, no conclusions can be reached concerning the degree of contribution of the instructional factor and the major field.

Astin's study also offers a possible explanation for the lack of significant differences between the two groups of freshmen. According to Astin, and to Feldman and Newcomb (1970), those who enter public institutions share more common demographic and ability characteristics than those who enter private colleges. Therefore, it may be that students match themselves to a type of institution rather than to a major field within that institution. Research including demographic data, ability data, and RJI scores on freshmen entering a variety of types of institutions would be useful in exploring this relationship.

In a more general way, Astin's findings, coupled with the prior research on college student development and the results of this study, argue for research designs that go beyond the simple distinctions of sex, major, and year in college and take into account more carefully both the characteristics of the institutions and the degree of involvement in college life. The literature on college student development clearly implies that peer contact is most strongly associated with the development; it is imperative that future research into Reflective Judgment take into account peer contact variables. Researchers into the Perry scheme (Widick, Knefelkamp, & Parker, 1975; Widick & Simpson, 1978; Mason, 1978; Touchton et al., 1978) have

consistently reported significant increases in Perry position when courses are designed with the stimulation of intellectual development as a primary purpose. Studies of the influences of specific classroom experiences and modes of instruction on RJI scores ought to be conducted. It is possible that the differences in impact between large and small institutions may be a function of more personalized instruction as well as increased involvement in the campus life and peer contact.

In short, while any conclusions based on these results are subject to the reservations cited above, the data are consistent with the interpretation that older students who persist in higher education show greater capacity to make reasoned judgments about intellectual problems than their younger, less educated counterparts, regardless of major field of study.

Nevertheless, even though the scores of seniors differed significantly from those of the freshmen on the Reflective Judgment Interview, seniors in this sample were still using whim as often as evidence and logic and could not see any way to consistently evaluate points of view as better or worse aside from individual idiosyncracies. In other words, these seniors were only somewhat better equipped than the freshmen to cope with the complex problems of modern society and were certainly much less developed intellectually than those responsible for their education would hope.

Furthermore, despite the common view that college students in technical fields tend to be more dogmatic in their thinking, no significant differences between those in engineering majors and those in humanities/social science majors were found in this study. The size of the difference in RJI scores between freshmen and seniors in humanities/social science was approximately the same as the difference between freshmen and seniors in engineering.

In summary, freshmen and seniors differed significantly on their Reflective Judgment Interview scores, but the difference in reasoning between freshmen and seniors in social science/humanities majors was approximately the same size as the difference between freshmen and seniors in engineering majors. While any conclusions based on these results must be subject to the reservations cited above, neverthless, the data are consistent with the following ideas. Older students completing higher education exhibit a greater awareness of multiple perspectives on the complex issued posed in the Reflective Judgment Interview. Despite the common conception that technical students tend to be more dogmatic, the engineering students in the present study did not differ from the social science/humanities students in the complexity with which they viewed the issues. While the seniors in the sample seemed more advanced than the freshmen in their capacity to make judgments about complex intellectual subjects, they appeared less developed intellectually than those responsible for their education would hope. Seniors used whim as often as evidence and logic in deriving their views, and they still could not see any way to consistently evaluate points of view aside from individual idiosyncracies.

Study III

The primary goal of this study is to assess the impact of graduate education on cognitive and cognitive-social developmental measures. To successfully address this question we must use a procedure which can isolate not only the effects associated with education but also those effects associated with competing explanations for group differences. Two common alternative explanations for observed growth are maturation and selection.

Our procedure specifies a priori patterns of possible results which best argue for one explanation over the other competing views. These patterns are then compared with the observed pattern to determine the best match.

When used with our population and purposes, this procedure specifies at minimum four groups of subjects. Two groups of graduate students at different stages in their training, and two groups of college graduates similar to the graduate groups on age, sex and verbal ability.

These four groups can be thought of as cells in a two by two matrix, the rows corresponding to enrollment status, and the columns to age. The patterns for the three competing effects, education, maturation and selection follow directly from this matrix (see Figure 3). If maturation is a factor the row means should be significantly different (Figure 3a). If selection is a factor the column means will differ significantly (Figure 3b). If education is a factor then an interaction between the column and rows should occur (Figure 3c). Further, this interaction should be due to a mean difference between the enrolled groups that is greater than the corresponding differences in the non-student groups.

The pattern we expect to see in the data will include a combination of the three pure forms shown in Figure 3. Although all these pure forms may be found in the data, the crucial test of the effects associated with education remains the row by column interaction described above.

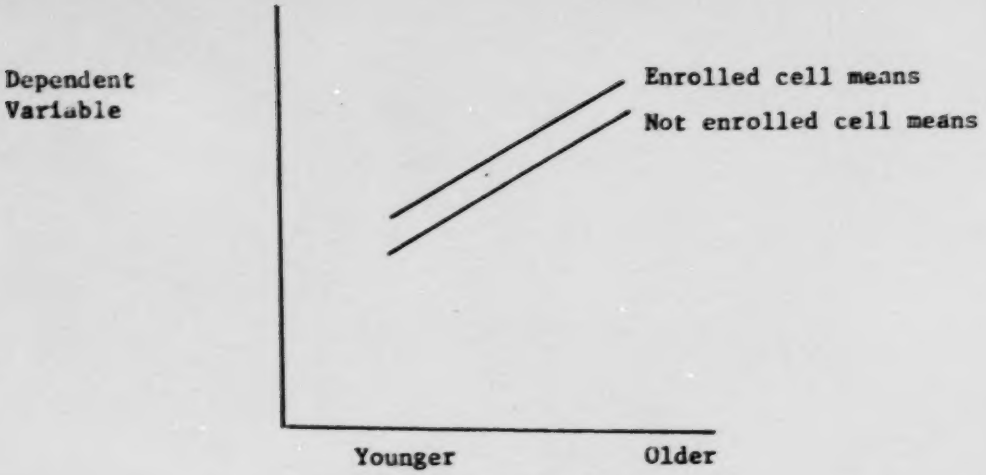

Figure 3a. Maturation Effect

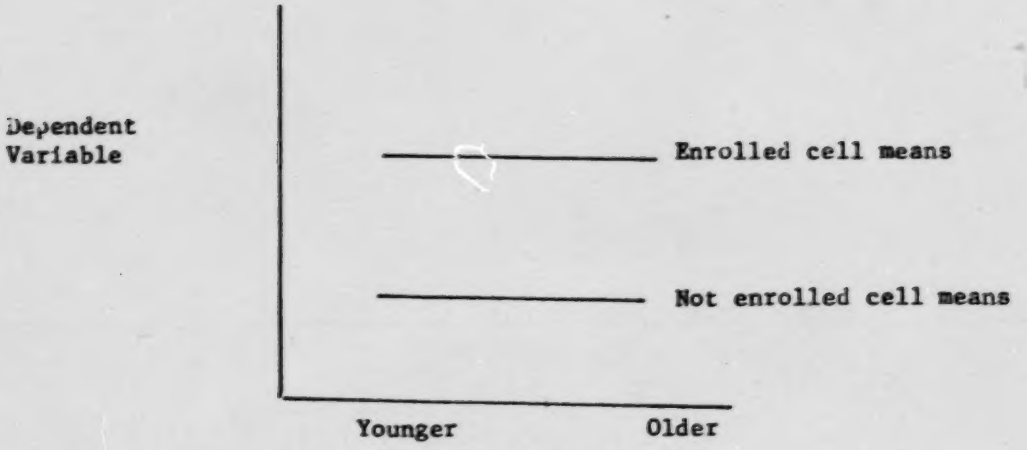

Figure 3b. Selection Effect

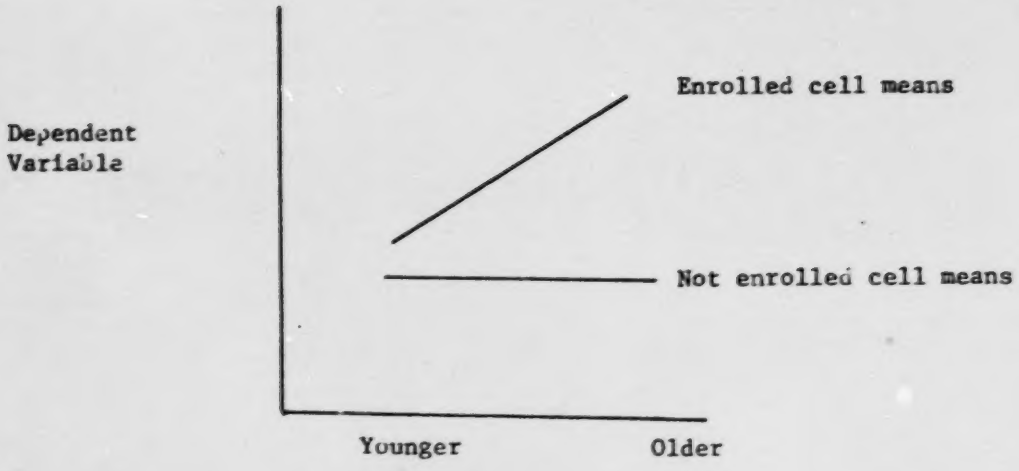

Figure 3c. Education Effect

Figure 3. Pure forms of the A priori patterns.

Method

Subjects

The sample collected for this study included eighty subjects, twenty advanced graduate students, twenty first-year graduate students, and forty college graduates who had not entered graduate school. These groups are equally divided by sex. All of the graduate students were enrolled at the University of Minnesota during the academic year 1978-79. The degree programs represented by these subjects were the traditional liberal arts majors excluding psychology.

Advanced graduate standing was defined as having completed the written preliminary exams required by the individual degree programs. All subjects in this group had been enrolled in their programs for at least three years.

The first year graduate subjects were all in their first year of graduate school; transfer students were not included in this sample.

The forty non-enrolled subjects had never enrolled in a graduate program but were of comparable age, and scholastic aptitude to the graduate subjects. In addition these subjects were currently living and working in the Twin Cities metropolitan area.

The selection process was as follows: a random sample of names was generated from lists of advanced graduate students supplied by the graduate school. These potential subjects were then contacted by mail or phone. This procedure was repeated until both ten males and females agreed to participate in the study.

First year subjects were selected in a similar fashion with the one exception that each subject had to have Scholastic Aptitude Test scores on file at the university that roughly matched the scores obtained from the advanced

group.

The non-enrolled group was recruited from newspaper advertisements, nominations from other subjects, and from lists supplied by the University of Minnesota Alumni Office. Of the potential subjects who responded to these recruitment procedures only those people who fit the age and test score profiles of the graduate group were included. These standards had to be relaxed for one sub cell due to our inability to locate 10 male college graduates who fit the profiles of the first year graduate males. The four males who were finally selected had scores that placed them in the upper third of their norm group. The rest of the sample had scores in the upper twenty-five percent of their norm group.

Instruments

Five instruments, the Reflective Judgment Interviews (RJI), Rest's Defining Issues Test (DIT), Loevinger's Sentence Completion Test (SCT), a general information questionnaire and Terman's Concept Mastery Test (CMT) were administered to each subject.

Reflective Judgment Interviews

The RJI (King, 1979; Kitchener, 1979) is a semi-structured interview consisting of four dilemmas, the origin of the Egyptian Pyramids, Objectivity of News Reporting, Creation versus Evolution, and the safety of chemical additives in food. The interview is guided by a standardized series of probes designed to elicit the subject's reaction to these dilemmas and his/her use of evidence to support their viewpoint. Each dilemma is defined for subjects by presenting two contradictory positions within the above content domains. The interview lasts for approximately 45 to 60 minutes.

The scoring procedure for the RJI utilizes trained raters who independently score each response to each dilemma. The individual dilemmas are summarized by three ratings which represent a major and two minor stage scores; for example a protocol scored 443 would indicate that the dominant stage for this subject is Stage 4 along with some evidence of Stage 3. These four triads, one set for each dilemma, are summed and the mean response obtained. This figure is used as the composite score for the subsequent analyses.

The DIT is an objective measure of moral reasoning development (Rest, 1979, Note 12). Subjects are required to read six stories, each describing a situation with competing social claims. Subjects are then asked to judge what he/she would do in the hypothetical situation and then rate and rank 12 issues relating to their decision. The majority of the 12 issues are stage typed to Kohlberg's (1969) stages of moral development. The remaining issues are designed as validity checks.

Subjects' ratings and rankings are summarized by the P and D scores. P scores utilize the ranking data and are defined as the weighted sum of the ranked principles types issues (Stages 5, 6). Percent P is the most common index found in the literature.

The D score takes into account the full range of rating responses and represents the use of principled thinking in relation to preconventional and conventional thinking.

The SCT of ego development (Loevinger, et al., 1970; Loevinger, Wessler, & Redmore, 1970) has two forms, one designed for men and the other for women. Both forms have 36 items. Each item requires the subject to complete a sentence stem.

To arrive at the composite score for a subject, trained raters first rate

(individually) all 36 items for ego level and then compare the cumulative frequency distribution of item rates to published tables of potential values. This procedure yields a single rating which is the measure of ego development used in the analyses.

The CMT (Terman, 1973) is an objective test of verbal reasoning ability. This test was designed to discriminate people who are superior in verbal reasoning and will therefore be appropriate for the present group of subjects. The CMT requires the subject to answer 115 synonyms and antonyms, plus 95 analogies. This test is primarily used in this study to statistically determine whether any group differences found by the primary measures mentioned above, can be attributed to differences in verbal reasoning.

This final instrument is included to provide an estimate of socio-economic status (SES) using Hollingsheads (Note 13) procedure. This estimate will be used to assess the relationship between SES and the primary measures in a manner similar to that described with the CMT.

Procedure

The length of the Reflective Judgment interview made it necessary to have subjects complete some of the measures without supervision. It was therefore determined that the DIT and SCT would be sent to each subject along with a covering letter confirming the interview date. The letter also requested that the two measures be completed before the interview.

On arrival at the testing location each subject was informed of the experimental procedures and a final consent form obtained. The research assistant then proceeded directly with the RJI. Following the interview the general information questionnaire was completed. The subject was then

escorted to the CMT testing room and asked to carefully read the instructions in the test booklet and begin the test. No time limit was enforced for any of the above procedures.

After the testing, each subject was debriefed and the measures completed at home, collected. Each subject was paid $10 upon completion of the instruments.

Results

Group Characteristics: Table 15 shows the obtained group profiles for the four groups. Although the SCT and DIT analyses used fewer subjects, due to rejected protocols, the group profiles do not significantly deviate from those shown in Table 15.

Reliability of the Reflective Judgment Interview, Interrater reliability and agreement, was computed on a random 25% of the protocols rated by both raters. These coefficients were .75 and .69 respectively, which are comparable to the results of other recent studies (Welfel, 1979; Kitchener, 1979). A further indication of the reliability of the measure was the internal consistency coefficient. Cronbach's alpha for all groups across all dilemmas was .79; again this figure is similar to those reported by other studies (Strange, 1978; Kitchener, 1979; Welfel, 1979).

What follows are the results for the primary measures. Inferences concerning the three hypothesized effects will be similar for each primary measure: if there is an additive effect of selection, one would expect the means for the enrolled and nonenrolled groups to differ significantly; if there is an additive effect of maturation, the main effect for age will be significant; if education is associated with the growth process, then differences between the two graduate groups will be greater than the differences between the

Table 15

Group Characteristics

Median verbal standardized test percentiles by group

	Younger	Older
Enrolled	mdn = 93 range = 77-99 N = 19	mdn = 95 range = 87-99 N = 20
Nonenrolled	mdn = 88 range = 63-98 N = 20	mdn = 90 range = 77-99 N = 20

Age

	Younger	Older
Enrolled	$\bar{x}$ = 23.85 sd = 1.88 N = 20	$\bar{x}$ = 30.30 sd = 4.97 N = 20
Nonenrolled	$\bar{x}$ = 25.15 sd = 2.48 N = 20	$\bar{x}$ = 30.05 sd = 2.06 N = 20

nonenrolled groups. This interaction is of crucial importance to the hypothesis of this study.

RJI: Table 16 presents the cell means and standard deviations for the total sample. These cells were compared using a hierarchical step down ANOVA procedure where the main effects are entered first followed by the interaction effects. When covariates are added, they are entered before the main effects.

The result of the RJI analysis shows that the three main effects, sex ($p < .001$), maturation ($p < .05$), and selection ($p < .01$) are all significant. All of the interactions, including the crucial interaction suggesting that education is associated with Reflective Judgment, are not significant ($p > .05$).

When SES and current verbal abilities measures are added to the analyses as covariates, the sex ($p < .01$) and selection ($p < .01$) effects remain; however the main effect for maturation is no longer significant ($p < .20$).

In sum, these results suggest that men score higher than women on the RJI and graduate students scored higher than nonstudents; a pattern which suggests a selection explanation for the differences between the students and nonstudents.

SCT: 78 valid protocols out of a possible 80 were rated by a trained rater, blind to conditions. To perform the following analyses Loevinger stages were assigned the numerical values listed in Table 17. Using these values Table 18 presents the means and standard deviations of the four subgroups. The ANOVA procedure described above was used to compare these groups. Although the means in Table 18 display the

Table 16

Cell Means and Standard Deviations for Reflective Judgment Scores by Group

	Younger	Older
Enrolled	$\bar{x}$ = 4.60 sd = .71 n = 20	$\bar{x}$ = 4.94 sd = .81 n = 20
Nonenrolled	$\bar{x}$ = 4.02 sd = .67 n = 20	$\bar{x}$ = 4.25 sd = .51 n = 20

Table 17

Loevinger Stage	2	2/D	D	D/3	3	3/4	4	4/5	5	6
Numerical Equivalent	1	1.5	2	2.5	3	3.5	4	4.5	5	6

hypothesized pattern which would suggest an association between education and ego development, all main and interaction effects are nonsignificant.

DIT: Sixty-seven subjects completed DIT's which passed both validity checks built into the scoring procedure.

Tables 19 and 20 present summary tables for both the P and D indices. The above mentioned ANOVA procedure was used to compare these cells. The main effects for age (P score, $p < .05$; D score $p < .01$); and enrollment (P score $p < .05$; D score $p < .05$) are significant for both indices. The sex effect is significant only with the D score (D score $p < .05$). The crucial interaction, suggesting an educational impact on moral reasoning, along with all other interactions are not significant. All of the main effects remain significant when current verbal ability and SES are entered into the analyses as covariates.

These results suggest that women score higher on the DIT than men when one considers the total distribution of item responses. Most importantly, these data show older subjects scoring higher than the younger subjects and enrolled subjects scoring higher than the nonenrolled subjects; findings which correspond to the a priori patterns representing selection and maturation. In short, one need not postulate an additive effect of education to account for the difference found in Table 19.

Table 18

Comparison of the Normal Ojive Score

of Enrolled and Nonenrolled Groups

	Younger	Older
Enrolled	$\bar{x}$ = 3.789 sd = .346 N = 19	$\bar{x}$ = 3.947 sd = .405 N = 19
Nonenrolled	$\bar{x}$ = 3.800 sd = .548 N = 20	$\bar{x}$ = 3.825 sd = .438 N = 20
ANOVA	Age effects F(1,74) = .805 NS Enrollment effect F(1,74) = .312 NS Age x Enrollment P(1,74) = .441 NS	

Table 19

Comparison on Z P Score of Enrolled and Nonenrolled Groups

	Younger		Older	
Enrolled	$\bar{x}$ = 62.37 sd = 11.22 N = 7	$\bar{x}$ = 64.59 sd = 10.38 N = 8	$\bar{x}$ = 65.18 sd = 10.94 N = 9	$\bar{x}$ = 65.74 sd = 11.40 N = 9
	Total Cell $\bar{x}$ = 63.55 sd = 10.45 N = 0		Total Cell $\bar{x}$ = 65.461 sd = 10.84 N = 18	
Nonenrolled	$\bar{x}$ = 51.48 sd = 21.26 N = 9	$\bar{x}$ = 52.98 sd = 8.55 N = 9	$\bar{x}$ = 56.86 sd = 10.86 N = 8	$\bar{x}$ = 67.88 sd = 7.78 N = 8
	Total Cell $\bar{x}$ = 52.23 sd = 15.74 N = 18		Total Cell $\bar{x}$ = 62.37 sd = 10.75 N = 16	

ANOVA:	Age effects	F(1,59) = 4.849	P .05
	Enrollment effect	F(1,59) = 5.59	P .05
	Sex effect	F(1,59) = 1.550	NS
	Age x Enrollment	F(1,59) = 1.67	NS
	All other interactions	NS	

Table 20

Comparison on D Score of Enrolled and Nonenrolled Groups

	Younger		Older	
Enrolled	$\bar{x}$ = 32.52 sd = 5.25 N = 7	$\bar{x}$ = 32.49 sd = 7.21 N = 8	$\bar{x}$ = 33.26 sd = 6.81 N = 9	$\bar{x}$ = 36.74 sd = 11.40 N = 9
	Total Cell $\bar{x}$ = 32.51 sd = 6.15 N = 15.0		Total Cell $\bar{x}$ = 34.83 sd = 6.33 N = 18	
Nonenrolled	$\bar{x}$ = 24.90 sd = 6.18 N = 9	$\bar{x}$ = 30.11 sd = 4.55 N = 9	$\bar{x}$ = 29.91 sd = 5.56 N = 8	$\bar{x}$ = 35.55 sd = 6.94 N = 8
	Total Cell $\bar{x}$ = 27.51 sd = 5.91 N = 18		Total Cell $\bar{x}$ = 32.73 sd = 6.73 N = 16	

ANOVA:	Age effects:	$F(1,59) = 7.434$ $p < .01$
	Enrollment effects:	$F(1,59) = 5.568$ $p < .05$
	Sex effect:	$F(1,59) = 5.865$ $p < .05$
	Age by enrollment	$F(1,59) = .928$ NS
	All other interactions	NS

Dissemination

Our attempts to disseminate the results of this project have taken three forms: journal articles or books, technical reports, and addresses to various forums by the members of the project staff.

Addresses

One of the first addresses based on this project was in fact a series of three addresses given to the Association of College Unions International in Minneapolis, Minnesota, March, 1980. James R. Rest delivered a paper entitled Moral judgment development in higher education, Mark L. Davison delivered a paper entitled How students tackle the tough questions, and Jane Lawson delivered a paper entitled Rearranging the living room.

In October of 1979, James R. Rest addressed the Menninger Foundation with a talk entitled The impact of higher education on moral judgment. Stephen Thoma and Mark Davison prepared a paper entitled Graduate education and moral judgment development for the Iowa Educational Research and Evaluation convention, Iowa City, Iowa, December 4-5, 1980. And, finally, James R. Rest recently made several presentations to the American Educational Research Association and the Society for Research in Child Development which were wholly or partly based on the research in this project; The impact of higher education on moral judgment development; Moral dilemmas of young adults; The major psychological components of morality; and Action advocacy in hypothetical moral dilemmas.

Technical Reports

Eight technical reports have been prepared in this project. The first, The impact of higher education on moral judgment development by James R. Rest

is an extension of his address to the Menninger Foundation mentioned above. The second report, by Mark L. Davison, Patricia King, Karen Kitchener, and Clyde A. Parker entitled The stage sequence concept in cognitive and social development subsequently appeared in the journal Developmental Psychology. Elizabeth R. Welfel prepared the third report, Reflective judgment and its relationship to academic field. Project report #4, Moral reasoning and college experience, was prepared by Joseph M. Volker.

James R. Rest prepared the fifth report, Moral dilemmas of young adults. Janet A. Schmidt and Mark L. Davison wrote Does college matter? Reflective judgment: How to tackle the tough questions, as the sixth report, a report which appeared in Moral Education Forum at the request of the editor. Stephen J. Thoma and Mark L. Davison prepared Graduate education, moral judgment development, and ego development as an outgrowth of a presentation to the Iowa Educational Research and Evaluation convention. And finally, this report constitutes the eighth and last in our series of project reports.

Journals and Books

A segment of James Rest's recent book, Development in judging moral issues is devoted to the relationship between education and moral judgment development. Two of the technical reports have appeared as articles. The second report, "The stage sequence concept in cognitive and social development" appeared in Developmental Psychology, 1980, 16, 121-131. The sixth report, "Does college matter? Reflective judgment: How to tackle the tough questions," appeared in Moral Education Forum, 1981 6 (1), 2-14. Finally, Clyde A. Parker authored an article in the Lyceum, a publication of the University of Minnesota College of Education. The article was entitled

"Teaching students to cope with uncertainty." The article focused on a number of student development and faculty development projects now going on at the University of Minnesota, including our project, "Higher education and cognitive-social development."

Complete citations for the addresses, technical reports, and publications are given in appendix B. These reflect our continuing effort to disseminate the results of the project.

Conclusions

Each of the three studies above summarizes the unique conclusions that follow from it. Rather than recapitulate those conclusions, this section will attempt to integrate them into a picture of student development consistent not only with the data from our own project, but other projects as well. Graduate and undergraduate education will be considered separately.

The studies on which our conclusions are based, including our own studies, suffer from methodological flaws. What we are trying to do in this section is briefly describe the trends in these less than perfect studies and the conclusions supported by the trends. Readers will no doubt keep in mind that the confidence one can place in the trends is limited by the imperfect data base on which they rest. In this section, our focus will be on the trends, not the methodological shortcomings of the studies.

Our conclusions refer only to the variables of this project. For instance, our conclusions about moral development refer only to moral reasoning as measured by the Defining Issues Test. Moral reasoning is only one aspect of ethical development. Likewise, Reflective Judgment is only one aspect of cognitive development and ego development as measured by Loevinger's instrument is only one aspect of personality development.

Graduate education

The studies to date have consistently found differences in Reflective Judgment Scores between advanced graduate students and beginning graduate students (or college students). This difference remains even after controlling for differences (or matching on) verbal ability. This finding is consistent with the idea that development in reasoning about complex issues having no simple right or wrong answer does occur in graduate school. These studies are all cross-sectional.

Do graduate students change in their reasoning at a faster rate than those not in graduate school? The third study in this report is the only one, to our knowledge, to address this question. The data offered an ambiguous answer. The difference between advanced and beginning graduate students was greater than the corresponding difference between control groups of nonattenders of graduate school, but not significantly so.

In the area of moral reasoning, studies again have consistently found that advanced graduate students have higher scores than beginning graduate students or college seniors. We found in study 3 that the difference between advanced and beginning graduate students was no greater than the difference between nonattending control groups. The result is consistent with the notion that those who do and do not attend graduate school after college develop in moral reasoning at about the same rate.

As for ego development, there is a growing body of evidence (Loevinger, 1979), including our own study 3, to suggest that there is little change in scores on Loevinger's measure beyond high school either for those in or out of school. Hence, there is little to suggest that a graduate education

promotes ego development as measured by Loevinger's test.

Undergraduate education

The literature tends to show that seniors have higher Reflective Judgment scores than freshmen. The studies are all cross-sectional, however. While we had hypothesized that liberal arts majors should promote more growth than "narrower" technical majors, our second study did not support this hypothesis. The freshman-senior difference was as wide in engineering as in the liberal arts. Do people attending college seem to change more than those who do not? There is to date only one study addressing this issue (Strange, 1978). It suggests that they do. Strange found that people differing in age and education by four years, but not those differing only in age, had significantly different Reflective Judgment scores. This study deserves replication with a longitudinal design.

While the Strange study is certainly good news for educators, the cognitive complexity of graduating seniors may not be what colleges would hope. Senior scores have typically fallen in the middle range where responses are characterized by (a) little evaluation of alternative views, (b) a tendency to treat all opinions as equally good, (c) a tendency to use whim and unsubstantiated belief in forming an opinion, and (d) a hesitancy to take a stance.

Study 1 not only confirms the finding that people in college grow in moral reasoning, it suggests they change faster than high school seniors who do not attend college. Contrary to the common belief in college student personnel work that the largest student changes occur in the first two years, the college attenders changed more than nonattenders.

in the third and fourth years beyond high school, but not in the first and second years. The vexing problem, however, remains one of clarifying how education has its effects on moral reasoning.

As for ego development, there is relatively little evidence to suggest that seniors differ significantly from freshmen. Hence, there is little evidence to suggest that undergraduate education has any impact on ego development as measured by Loevinger's instrument.

At least at the undergraduate level, the evidence to date is consistent with the conclusion that higher education does have an impact on students' reasoning about moral issues as measured by the Defining Issues Test and students' reasoning about complex intellectual issues which have no simple right or wrong answer as measured by the Reflective Judgment Interview. How those effects occur is not clear. Until we better understand which academic and nonacademic experiences mediate the educational effects, it will be difficult to design curricula specifically to promote those effects.

Appendix

Addresses, Technical Reports and Publications

Davison, M. L. How students tackle the tough questions. Paper presented to the Association of College Unions International, Minneapolis, Minnesota, March, 1980.

Davison, M. L., King, P. M., Kitchener, K. S. & Parker, C. A. The stage sequence concept in cognitive and social development. (Technical report #2, Higher education and cognitive-social development project.) Minneapolis, Minnesota: University of Minnesota, 1979.

Davison, M. L., King, P. M., Kitchener, K. S., & Parker, C. A. The stage sequence concept in cognitive and social development. Developmental Psychology, 1980, 16, 121-131.

Parker, C. A. Teaching students to cope with complexity and uncertainty. Lyceum, 1980, 3 (1), 2-6.

Rest, J. R. The impact of higher education on moral judgment development. Paper presented to the Menninger Foundation, Topeka, Kansas, October 13-14, 1979.

Rest, J. R. The impact of higher education on moral judgment development (Technical report #1, Higher education and cognitive-social development project). Minneapolis, Minnesota: University of Minnesota, 1979.

Rest, J. R. Development in judging moral issues. Minneapolis, Minnesota: University of Minnesota Press, 1979.

Rest, J. R. Moral judgment development in higher education. Paper presented to the Association of College Unions International, Minneapolis, Minnesota, March, 1980.

Rest, J. R. Moral dilemmas of young adults (Project report #6, "Higher education and cognitive-social development project"). Minneapolis, Minnesota: University of Minnesota, 1980.

Rest, J. R. The impact of higher education on moral judgment development. Presented at the AERA Convention, April 14, 1981, Los Angeles.

Rest, J. R. Action advocacy in hypothetical moral dilemmas. Presented at the SRCD Convention, April 3, 1981, Boston.

Schmidt, J. A., & Davison, M. L. Does college matter. Reflective judgment: How students tackle the tough questions. (Project report #6, Higher education and cognitive-social development project). Minneapolis, Minnesota: University of Minnesota, 1980.

Schmidt, J. A., & Davison, M. L. Does college matter? Reflective judgment: How to tackle the tough questions. Moral Education Forum, 1981, 6 (1), 2-14.

Thoma, S. J., & Davison, M. L. Graduate education and moral judgment development. Paper presented to the Iowa Educational Research and Evaluation Convention, Iowa City, Iowa, December 4-5, 1980.

Thoma, S. J., & Davison, M. L. Graduate education, ego development and moral judgment development (Project report #7, Higher education and cognitive-social development project). Minneapolis, Minnesota: University of Minnesota, 1981.

Volker, J. M. Moral reasoning and college experience (Project report #4, Higher education and cognitive-social development project). Minneapolis, Minnesota: University of Minnesota, 1979.

Welfel, E. R. Reflective judgment and its relationship to academic field (Project report #3, Higher education and cognitive-social development project). Minneapolis, Minnesota: University of Minnesota, 1979.

Reference Notes

1. Rest, J. R. The stage concept in moral judgment research. Unpublished manuscript, University of Minnesota, 1977.
2. Davison, M. L., & Robbins, S. The reliability and validity of objective indices of moral development. Unpublished manuscript, University of Minnesota, 1977.
3. King, P. M., Kitchener, K., Parker, C. A., & Davison, M. L. Moral development, ego development, and reflective judgment. Manuscript in preparation.
4. Erickson, V. L. Personal communications, 1977.
5. Rest, J. R. Moral judgment related to sample characteristics. (Tech. Rep. #8703 MH 24988). Minneapolis, Minnesota: University of Minnesota, 1976.
6. Lockwood, A. L. The effects of values clarification and moral development curricula on school-age subjects: A critical review of recent research. Unpublished manuscript, University of Wisconsin, 1977.
7. Lawrence, J. A. Moral judgment intervention studies using the Defining Issues Test. Paper presented at the meeting of the Society for Research in Child Development, New Orleans, March, 1977.
8. Clinchy, B., & Zimmerman, C. Personal communication, 1977.
9. Blake, L. A measure of developmental change: A cross-sectional study. Paper presented at the Annual Meeting of the American Psychological Association, Washington, D. C., 1976.
10. Widick, C., & Simpson, D. The use of developmental concepts in college instruction. Paper presented at the student Development Conference, Minneapolis, May, 1976.

11. Rest, J. R. Manual for the Defining Issues Test. Unpublished manuscript, University of Minnesota, 1974.

12. Rest, J. R. Revised manual for the Defining Issues Test. Minneapolis, Minnesota: Minnesota Moral Research Project (330 Burton Hall, 178 Pillsbury Drive S. E., Minneapolis, Minnesota 55455), 1979.

13. Hollingshed, A. B. Two factor index of social position. Unpublished manuscript, Yale Station, New Haven, Connecticut, 1965.

References

Astin, A. Four critical years: Effects of college on beliefs, attitudes, and knowledge. San Francisco: Jossey-Bass, 1978.

Bereiner, J. E., Jr. A psychological education intervention for teacher development. Unpublished doctoral dissertation, University of Minnesota, 1976.

Blake, L. A measure of developmental change: A cross-sectional study. Paper presented at the Annual Meeting of the American Psychological Association, Washington, D. C., 1976.

Blasi, A. A developmental approach to responsibility training. Unpublished doctoral dissertation, Washington University, 1971.

Boyd, D. Some thoughts on a comparison of Perry and Kohlberg. Unpublished manuscript, Harvard University, 1972.

Brabeck, M. The relationship between critical thinking skills and the development of reflective judgment. Unpublished manuscript, University of Minnesota, 1978.

Broadhurst, B. P. Report: Defining Issues Test. Unpublished manuscript, Colorado State University, 1980.

Broughton, J. The development of natural epistemology in adolescence and early adulthood. Unpublished doctoral dissertation, Harvard University, 1975.

Campbell, T., & Stanley, C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Candee, D. Ego developmental aspects of new left ideology. Journal of Personality and Social Psychology, 1974, 30, 620-630.

Clinchy, B., & Zimmerman, C. Cognitive development in college. Unpublished manuscript, Wellesley College, 1975.

Clinchy, B., Lief, J., & Young, P. Epistemological and moral development in girls from a traditional and progressive high school. Journal of Educational Psychology, 1977, 69, 337-343.

Coder, R. Moral judgment in adults. Unpublished doctoral dissertation, University of Minnesota, 1975.

Crowder, J. W. The Defining Issues Test and correlates of moral judgment. Unpublished master's thesis, University of Maryland, 1976.

Davison, M. L. On a metric, unidimensional unfolding model for attitudinal and developmental data. Psychometrika, 1977, 42, 523-548.

Davison, M. L., King, P. M., Kitchener, K. S., & Parker, C. A. The stage sequence concept in cognitive and social development. Developmental Psychology, 1980, 16, 121-131.

Davison, M. L., Parker, C. A., & Kitchener, K. Higher education and cognitive-social development. Unpublished manuscript, University of Minnesota, 1978.

Davison, M. L., & Robbins, S. The reliability and validity of objective indices of moral development. Applied Psychological Measurement, 1978, 2 (3), 389-401.

Davison, M. L., Robbins, S., & Swanson, D. B. State structure in objective moral judgments. Developmental Psychology, 1978, 14, 137-146.

Dewey, J. How we think: A restatement of the relation of reflective thinking to the educative provess. Lexington: D. A. Heath and Company, 1933.

Dortzbach, J. R. Moral judgment and perceived locus of control: A cross-sectional developmental study of adults, aged 25-74. Unpublished doctoral dissertation, University of Oregon, 1975.

Erickson, V. L. Psychological growth for women: A cognitive-developmental curriculum intervention. Counseling and Values, 1974, 18, 102-116.

Erickson, V. L. Psychological growth for women: A cognitive-developmental curriculum intervention. Unpublished doctoral dissertation, University of Minnesota, 1973.

Exum, H. A. Cross-age and peer teaching: A deliberate psychological education curriculum for junior college students. Unpublished doctoral dissertation, University of Minnesota, 1977.

Feldman, K. A., & Newcomb, T. M. The impact of college on students. (Vols. I and II) San Francisco: Jossey-Bass, 1970.

Haan, N., Stroud, J., & Holstein, J. Moral and ego stages in relationship to ego processes: A study of "hippies." Journal of Personality, 1973, 41, 596-612.

Hauser, S. T. Loevinger's model and measure of ego development: A critical review. Psychological Bulletin, 1976, 83, 928-955.

Hays, W. L. Statistics. Chicago: Holt, Rinehart, and Winston, 1963.

Heffernan, J. Identity formation, identity orientations, and sex differences related to college environment features: A comparative study of conventional and innovative undergraduate programs. Unpublished doctoral dissertation, University of Michigan, 1971.

Hollingshead, A. B. Two factor index of social position. Unpublished manuscript. Yale Station, New Haven, Connecticut, 1965.

Holstein, C. B. Irreversible step wise sequence in the development of moral judgment: A longitudinal study of males and females. Child Development, 1976, 47, 51-61.

Hoppe, C. F. Ego development and conformity behaviors (Doctoral dissertation, Washington University, 1972). Dissertation Abstracts International, 1972, 33, 6060B.

Hurt, B. L. Psychological education for college students: A cognitive developmental curriculum (Doctoral dissertation, University of Minnesota, 1974). Dissertation Abstracts International, 1975, 35, 5119A.

Kaseman, T. C. A longitudinal study of moral development of the West Point class of 1981. West Point, New York: Department of Behavioral Sciences and Leadership, United States Military Academy, 1980.

Keniston, K., & Gerzon, M. Human and social benefits. In Wilson, L., & Mills, O. (Eds.), Universal American Education. Washington, D.C.: American Council of Education, 1972.

King, P. M. The development of reflective judgment and formal operational thought in adolescents and young adults. Unpublished doctoral dissertation, University of Minnesota, 1977.

King, P. M. William Perry's theory of intellectual and ethical development. In Knefelkamp, K., Widick, C., & Parker, C. (Eds.), Applying new developmental findings. San Francisco: Jossey-Bass, 1978.

King, P. M., & Parker, C. Assessing intellectual development in the college years: A report of the instructional improvement project, 1976-1977. Unpublished manuscript, University of Minnesota, 1978.

Kitchener, K. The development and measurement of reflective judgment in late adolescents and young adults. Unpublished doctoral dissertation, University of Minnesota, 1977.

Kitchener, K. Personal communication, 1979.

Knefelkamp, L., & Sleptiza, R. A cognitive model of career development: An adaptation of the Perry scheme. In Parker, C. (Ed.), Encouraging development in college students. St. Paul, Minnesota: University of Minnesota Press, 1978.

Kohlberg, L. Stage and sequence: the cognitive-developmental approach to socialization. In Goslin, D. (Ed.), Handbook of socialization theory and research. Chicago: Rand McNally, 1969, 347-480.

Kohlberg, L. From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In Mischel, T. (Ed.), Cognitive development and epistemology. New York: Academic Press, 1971.

Kuhn, D. Short-term longitudinal evidence for the sequentiality of Kohlberg's early stages of moral development. Developmental Psychology, 1976, 12, 162-166.

Kuhn, T. S. Reflections on my critics. In Lakatos, I., & Musgrave, A. (Eds.), Criticism and the growth of knowledge. Cambridge, England: Cambridge University Press, 1970.

Kurfiss, J. Late adolescent development: A structural epistemological approach. Unpublished doctoral dissertation, University of Washington, 1975.

Lambert, H. Comparison of cognitive developmental theories of ego and moral development. Proceedings of the 80th Annual Convention of the American Psychological Association, 1972, 7, 115-116.

Lawlis, G. F., & Lu, E. Judgment of counseling process: Reliability, agreement, and error. Psychological Bulletin, 1972, 78, 17-20.

Lawrence, J. N. Moral judgment intervention studies using the Defining Issues Test. Journal of Moral Education, 1980.

Loevinger, J. Ego development: Conceptions and theories. San Francisco: Jossey-Bass, 1976.

Loevinger, J. Construct validity of the Sentence Completion Test of ego development. Applied Psychological Measurement, 3 (3), 1979, 281-311.

Loevinger, J., & Wessler, R. Measuring ego development: Construction and use of a sentence completion test. San Francisco: Jossey-Bass, 1970.

Loevinger, J., Wessler, R., & Redmore, C. Measuring ego development 2: Scoring manual for women and girls. San Francisco: Jossey-Bass, 1970.

Mason, K. Effects of developmental instruction on the development of cognitive complexity, locus of control, and empathy on beginning counseling graduate students. Unpublished master's thesis, University of Maryland, 1978.

Mentkowski, M. Evaluation of the Alverno College values education project. In preparation.

Meyer, P. Intellectual development of college students as measured by religious content. Unpublished doctoral dissertation, University of Minnesota, 1975.

Mines, R. A. An investigation of the developmental levels of reflective judgment and associated critical thinking skills. Unpublished manuscript, University of Iowa, 1979.

Perry, W. G., Jr. Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart, and Winston, 1970.

Piaget, J. The moral judgment of the child. (M. Gabain, trans.) New York: The Free Press, 1965. (Originally published, 1932).

Pohl, R., & Pervin, L. Academic performance as a function of task requirements and cognitive style. Psychological Reports, 1968, 22, 1017-1020.

Popper, K. Objective knowledge. Oxford, England: Oxford University Press, 1972.

Redmore, C., & Waldman, K. Reliability of a sentence completion measure of ego development. Journal of Personality Assessment, 1975, 39, 236-243.

Rest, G. J. Voting preference in the 1976 presidential election and the influences of moral reasoning. Unpublished manuscript, University of Michigan, 1977.

Rest, J. R. The hierarchical nature of moral judgment. Journal of Personality, 1973, 41, 86-109.

Rest, J. R. Longitudinal study of the Defining Issues Test: A strategy for analyzing developmental change. Developmental Psychology, 1975, 11, 738-748.

Rest, J. R. New approaches in the assessment of moral judgment. In Lickona, T. (Ed.), Moral development and behavior: Theory, research and social issues. New York: Holt, Rinehart and Winston, 1976.

Rest, J. Moral judgment related to sample characteristics. Final report to the National Institute of Mental Health, Grant #8703 MH 2498, Developing an objective test of moral judgment, Appendix A, 1976.

Rest, J. R. Development in judging moral issues. Minneapolis: University of Minnesota Press, 1979.

Rest, J. R., Cooper, D., Coder, R., Masanz, J., & Anderson, D. Judging the important issues in moral dilemmas--an objective test of development. Developmental Psychology, 1974, 10, 491-501.

Rest, J. R., Davison, M. L., & Robbins, S. Age trends in judging moral issues: a review of cross-sectional, longitudinal and sequential studies of the Defining Issues Test. Child Development, 1978, 49, 263-279.

Riegel, K. Dialectic operations: The final period of cognitive development. Human Development, 1973, 16, 346-370.

Schroeder, H., Driver, M., & Struefert, S. Human information processing. New York: Holt, Rinehart and Winston, 1967.

Scrivens, M. Maximizing the power of causal investigation: The modus operandi method. In Glass, G. V. (Ed.), Evaluation studies review Annual, Vol. 1. Beverly Hills, California: Sage Publications, 1976.

Sheehan, T. J., Husted, S. D., & Candee, D. The development of moral judgment over three years in a group of medical students. Paper presented at AERA Convention, Los Angeles, April, 1981.

Shoff, S. Reasoning in younger and older students and non-students. Unpublished manuscript, University of Utah, 1978.

Stephenson, B. W., & Hunt, C. Intellectual and ethical development: A dualistic curriculum intervention for college students. The Counseling Psychologist, 1977, 6 (4), 39-42.

Strange, C. Intellectual development, motive for education and learning styles during the college years: A comparison of adult and traditional age college students. Unpublished doctoral dissertation, University of Iowa, 1978.

Sullivan, E. V., McCullough, G., & Stager, M. A developmental study of the relationship between conceptual, ego, and moral development. Child Development, 1970, 41, 399-411.

Sullivan, P. J. A Curriculum for stimulating moral reasoning and ego development in adolescents (Doctoral dissertation, Boston University, 1975). Dissertation Abstracts International, 1975, 36, 1320A.

Terman, L. M. Concept mastery test manual. New York: The Psychological Corporation, 1973.

Touchton, J. G., Wertheimer, L. G., Cornfield, J. L., & Harrison, K. H. Career planning and decision-making: A developmental approach to the classroom. The Counseling Psychologist, 1977, 6 (4), 42-47.

Touchton, J. G., Wertheimer, L. G., Cornfield, J. L., & Harrison, K. H. Career planning and decision making. In Parker, C. (Ed.), Encouraging development in college students. St. Paul, Minnesota: University of Minnesota Press, 1978.

Welfel, E. A pilot study of the relationship between William Perry's scheme of intellectual development of college students and the Omnibus Personality Inventory. Unpublished master's paper, University of Minnesota, 1975.

Whiteley, J. Character development in college students. Schenectady, N. Y.: Character Education, in press.

Widick, C. C., Knefelkamp, L. L., & Parker, C. A. The counselor as developmental instructor. Counselor Education and Supervision, 1975, 14, 286-296.

Widick, C. C., & Simpson, D. Developmental concepts in college instruction. In Parker, C. (Ed.), Encouraging development in college students. St. Paul, Minnesota: University of Minnesota Press, 1978.

Wilson, W. W. The spirit of learning, 1909. As quoted in the Phi Beta Kappa Handbook for New Members, 1970, iv.